A Publication of the American Psychological Association

The Journal of Abnormal and Social Psychology

Founded by Morton Prince, 1906

CONTENTS

ARTICLES

Prediction of Flying Success by Clinical Analysis of Test Protocols 485
W. H. Holtzman and S. B. Sells

The Influence of Set on Semantic Generalization 493
Nathaniel M. Lipsett

Conditioning Level in the Behavior Disorders 497
Janet A. Taylor and Kenneth W. Spence

The Role of Expectancy in the Performance of Posthypnotic Behavior 503
Seymour Fisher

Self-Perception, Perceptual Defense, and Adjustment 508
Bernard Chodorkoff

An Experimental Investigation of the Effects of Unstable Interpersonal Relations in a Group 511
Leon Festinger and Harmon A. Kelley

Projection and Displacement: A Cross-Cultural Study of Folk-Tale Aggression 515
George O. Wright

A Movement Interpretation of Threat 520
Abbie W. Lundfield

The Color-Form Responses of Normal, Psychotic, and Neurotic Subjects 531
J. D. Keehn

Continued on inside front cover

Volume 49 OCTOBER 1954 Number 4, Part 1

CONTENTS—*Continued*

ARTICLES — PAGE

Anxiety and the Mode of Response . . . 538
Robert E. Silverman

An Analysis of the Cattell-Luborsky Humor Test into Homogeneous Scales . . . 543
James K. Yarnold and Marvin H. Berkeley

Some Effects of Unequal Distribution of Information upon Group Performance in Various Communication Nets . . . 547
Marvin E. Shaw

Some Effects of Unequal Distribution of Information in a Wheel Group Structure . . . 554
J. C. Gilchrist, Marvin E. Shaw, and L. C. Walker

The Effect of One Partner's Success in a Relevant Task on the Interaction of Observer Pairs . . . 557
Bernard Maucner

Some Personality Correlates of Stimulus Generalization under Stress . . . 561
Charles W. Eriksen

Tactual Perception of Form by the Brain-Injured . . . 566
Alan O. Ross

The Consistency of Interpersonal Behavior Judgments Made on the Basis of Short-Term Interactions in Three-Man Groups . . . 573
Robert R. Blake, Jane Srygley Mouton, and Benjamin Fruchter

CASE REPORT

A Blind Analysis of a Case of Multiple Personality Using the Semantic Differential . . . 579
Charles E. Osgood and Zella Luria

CRITIQUE AND NOTES

Effect of Increasing the Number of Rorschach Responses on *Sum C* and *M*: A Note on Fiske and Baughman's Study . . . 592
Conan Kornetsky and Donald L. Gerard

Adequacy of the Reporting-Test Measures for Detecting Reactions to Frustration . . . 593
W. Bruce Douglass and Kenneth B. Brown

The Stability of Autokinetic Judgments . . . 595
J. H. Rohrer, S. H. Baron, E. L. Hoffman, and D. V. Swander

REVIEWS . . . 598

NEW BOOKS RECEIVED . . . 605

AUTHOR INDEX . . . 607

Volume 49 OCTOBER 1954 Number 4, Part I

The Journal of Abnormal and Social Psychology

Editor: J. McV. Hunt, University of Illinois
Associate Editor: Arthur Jenness, Williams College
Book Review Editor: Stanley G. Estes, Harvard University
Managing Editor: Lorraine Bouthilet

CONSULTING EDITORS

Gordon W. Allport
Harvard University

Hadley Cantril
Princeton University

Lee J. Cronbach
University of Illinois

John Dollard
Yale University

Leonard W. Doob
Yale University

Leon Festinger
University of Minnesota

Charles M. Harsh
U.S. Navy Electronics Laboratory
San Diego

William A. Hunt
Northwestern University

W. Ernst Kris
Yale University School of Medicine

Donald W. MacKinnon
University of California

Quinn McNemar
Stanford University

O. Hobart Mowrer
University of Illinois

Lawrence I. O'Kelly
University of Illinois

Charles E. Osgood
University of Illinois

Saul Rosenzweig
Washington University

Harold Seashore
Psychological Corporation

David Shakow
National Institute of Mental Health

Richard L. Solomon
Harvard University

Ross Stagner
University of Illinois

Ivan Steiner
University of Illinois

Elizabeth Collins Wesman
White Plains, N. Y.

Robert W. White
Harvard University

Charles Wrigley
University of Illinois

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.
1333 SIXTEENTH STREET N.W., WASHINGTON 6, D.C.
AND MT. ROYAL & GUILFORD AVES., BALTIMORE 2, MD.

Entered as second-class matter February 15, 1954 at the Post Office at Baltimore, Maryland under the Act of February 28, 1925. Acceptance for mailing at special rate of postage provided for in Section 34.40 (e), P. L. and R. Additional entry at Washington, D.C.

Copyright 1954 by the American Psychological Association, Inc.

The JOURNAL is issued quarterly in January, April, July, and October. The yearly volume averages 544 pages. The subscription price per volume is $7.00; foreign, $7.50. Single issues are $2.00.

Original manuscripts and editorial communications should be sent to:

DR. J. MCV. HUNT
Department of Psychology
University of Illinois
Urbana, Ill.

Books for review should be sent to:

DR. STANLEY G. ESTES
Psychological Clinic
64 Plympton Street
Cambridge 38, Mass.

Fifty free offprints are supplied to contributors of articles. For book reviews, five copies of the JOURNAL are supplied gratis if requested. Authors of prior publication articles receive no gratis offprints or journals.

Printed and mailed by:

WAVERLY PRESS, INC.
Baltimore 2, Maryland

Subscriptions and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 Sixteenth Street N.W.
Washington 6, D. C.

Changes of address must reach the Subscription Office by the 25th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PREDICTION OF FLYING SUCCESS BY CLINICAL ANALYSIS OF TEST PROTOCOLS[1]

W. H. HOLTZMAN
University of Texas
AND S. B. SELLS
USAF School of Aviation Medicine

IN RELATION to the problem of developing a battery of personality and motivational measures to predict adjustment to military and combat flying, the following questions were considered: (*a*) Can a group of expert clinical psychologists, given a description of the problem and a definition of the criterion, predict this criterion on the basis of a battery of personality tests? (*b*) If such predictions can be made successfully, what are the effective predictor signs which the clinicians use, and how do they integrate these signs to obtain successful predictions? The experiment reported below was undertaken primarily for the purpose of developing hypotheses for the systematic, quantitative scoring of a group of experimental psychiatric screening tests for flying personnel.

The six tests in this study were part of one experimental battery of group tests similar to those used by psychologists in clinical practice. These tests are: (*a*) Background Information, a biographical history inventory; (*b*) Ink-Blot Test, a group adaptation of the Rorschach test; (*c*) Feeling and Doing, a psychosomatic inventory reporting both symptoms experienced and what the subject did about them; (*d*) What Is He Saying?—a sentence completion test designed for Air Force use; (*e*) L-D Test, a group test booklet of the Szondi test; and (*f*) Drawing Test, a group version of the Draw-A-Person test. These tests have been described in detail by Sells and others in a series of Air Force publications (1, 2, 4, 6, 7).

A major problem in adapting these clinical testing procedures to selection use is that of developing adequate formal, quantitative scoring techniques that meet the requirements of psychometric standards and at the same time incorporate the interpretive schemes of the clinical method. This problem was attacked in a number of ways, for the most part by empirical analyses in which empirical scoring keys were developed by painstaking and systematic comparison of the responses of discrete bipolar criterion groups and cross-validated on independent samples.[2] The present experiment was undertaken as one part of the coordinated attack on this problem.

METHOD

After careful consideration of the test and criterion data, the probable availability of skilled clinicians to evaluate the test protocols, and the objectives of the research program, an experimental design for the clinician study was developed. From a total of 1,504 test subjects (*Ss*) for whom complete criterion data were available, 100 aviation cadets were selected for the present study. Fifty of them had made a highly successful adjustment to the stress of flight training, and 50 had been eliminated from the program because they developed overt personality disturbances and were unable to complete the training course. The nature of the overt personality disturbances varied, but in every case the disturbance was sufficiently severe to be called to the attention of the flight surgeon, psychiatrist, and psychologist. The criteria used for selection of the 100 cadets are:

Successful Group

1. Passed the flight training course
2. Pilot aptitude of Stanine 6 or higher

[1] The research reported here was conducted under terms of Contract No. AF 33(038)-13887 between the U. S. Air Force and The University of Texas in coordination with the Department of Clinical Psychology, USAF School of Aviation Medicine, Randolph Field, Texas. This research was discussed at a symposium entitled "Clinical Prediction of Flying Success," American Psychological Association, Washington, D. C., September, 1952.

[2] Of the six tests used in this study, empirical studies have thus far developed valid scoring keys for three. Reports on the Rorschach (6) and Sentence Completion Test (7) have been published. Unpublished results on the Background Information inventory are reported in a general discussion of the psychiatric selection research program by Sells and Barry (5). Two studies of the Draw-A-Person test, one by Anastasi and Foley (1) and an unpublished study by Nurse and Barry, a series of analyses of the Szondi test by Ellis and Sells, and an extensive study of the Feeling and Doing test by Barry and Cobb have all yielded negative findings on cross-validation of experimental keys.

3. No evidence of personality or disciplinary difficulty
4. Leadership rating above the 75th percentile as judged by fellow cadets
5. Satisfactory ratings by four independent instructors on over-all suitability for role of Air Force pilot and officer

Unsuccessful Group

1. Elimination from the flight training program for reasons other than physical disability
2. Pilot aptitude of Stanine 6 or higher
3. Evidence of overt personality disturbance contributory to elimination

Since the number of cadets meeting these criteria was more than 50 in each group, 50 successful and 50 unsuccessful cadets were selected at random from the available cases. Code numbers were assigned to these 100 individuals, and five independent sets of 20 cadets each were drawn up randomly. In this manner, the a priori probability that a given cadet passed or failed was fixed at .50, with the number of successful cadets in a given set of 20 varying randomly about an expected value of 10. Consequently, the success of a given clinician could be evaluated directly by use of the binomial distribution.

Each test protocol was photographed, and sufficient photostatic copies were made to provide each participating psychologist with a complete set of test materials for his particular group of 20 cadets.

Meanwhile a select list of psychologists, many of whom had served as military psychologists during World War II, was compiled; it consisted chiefly of recognized authorities in clinical procedures as applied to test interpretation and psychodiagnosis. As partial remuneration for the time-consuming clinical evaluation of 20 cadets, a small honorarium of $100 was offered each collaborator for his participation. The 19 psychologists who accepted this invitation are:

Lawrence Abt	James Layman
Arthur Benton	Ivan Mensh
Sidney Bijou	Daniel Miller
Gerald Blum	Z. A. Piotrowski
Charles Cofer	Victor Raimy
Molly Harrower	Thomas Richards
Max Hutt	William Schofield
Seymour Klebanoff	Meyer Williams
Bruno Klopfer	Lee Winder
Samuel Kutash	

TABLE 1

EXPERIMENTAL DESIGN

Group	No. of Clinicians	No. of Cadets	No. of Cadets Who Passed
1	3	20	9
2	3	20	8
3	3	20	12
4	3	20	12
5	7	20	9
Total	19	100	50

To provide uniform background information concerning the nature of the criterion and the flight-training program, three chapters from Sells's (4) comprehensive outline of the research program which described the population and the case histories of unsuccessful cadets and outlined the critical requirements for success were included with the test materials. Detailed instructions for making and recording judgments, a brief summary of the over-all experimental design, and a short description of each test were also enclosed.

Two forms of recording sheet were used, one for recording judgments based on only one test at a time and one for recording global evaluations based on consideration of all or part of the available data for a given individual. Although each psychologist was encouraged to use first the single-test approach and then the global approach, the single-test analysis was made entirely optional. In addition to judging pass-fail for each cadet, each psychologist was asked to state, where possible, the cues which formed the basis for his evaluation. A three-point "confidence" scale was also provided so that each clinician could state the degree of confidence he had in the validity of each judgment he made. A confidence rating of "A" was to be given when the clinician felt quite sure of his judgment; a rating of "B" was to be used when he felt his judgment was questionable; and a rating of "C" denoted no confidence whatsoever in the judgment.

Table 1 summarizes the essential features of the experimental design. Four of the five sets of test protocols were each judged independently by three different psychologists; the fifth set was evaluated by seven psychologists. Replication through judges made possible a study of the clinician as a variable and permitted the use of pooled judgments for a single case. The additional replication in the fifth group was designed to provide a more accurate estimate of interclinician variability than was possible with only three clinicians per cadet.

The experimental design used in the present study permits several different approaches to testing the general hypothesis that adjustment to the training program could be predicted significantly better than chance by skilled clinicians from an examination of individual test protocols. Two sets of predictions were made by most of the psychologists. Every psychologist made a prediction of success or failure for each individual in his group of 20 cadets. One class of predictions consists of these over-all or global evaluations. In addition to using the global approach, 12 psychologists made predictions for each cadet based upon consideration of only one test at a time. A second class of predictions consists of these single-test evaluations. In either case, the data may be examined for one psychologist independently of the others, or combined predictions can be obtained by pooling the judgments of individual psychologists for each cadet.

RESULTS

Let us first examine the validity results of the global judgments. The number of correct judgments among the 19 psychologists ranged from 4 to 14 with a mean of 10.2. Since 20

TABLE 2

SUCCESS OF OVER-ALL (GLOBAL) EVALUATIONS FOR INDIVIDUAL CLINICIANS*

Group	Clinician Code No.†	Correct Predictions	Group	Clinician Code No.†	Correct Predictions
1	C	11	4	G	12
	A	10		Q	13
	B	9		F	12
2	R	4	5	I	8
	O	11		D	8
	M	8		E	10
3	L	14		K	9
	S	14		P	12
	J	13		H	10
				N	6

* Chance expectation is 10 correct out of 20. Mean number of correct predictions per clinician is 10.2.

† The code numbers have no relation to the order of listing of the clinicians' names.

cadets were judged by each psychologist, and since the a priori probability that a single prediction is correct by chance alone is .50, the expected chance number of correct predictions for a single clinician is 10. The results of this analysis are presented in Table 2. If we disregard confidence ratings, none of the psychologists was able to predict adjustment to training significantly better than chance with the global approach to test evaluation.[3]

A similar analysis was made using only the predictions with confidence ratings of A, which signified that the clinician was quite sure of his judgment. The number of over-all predictions that were rated A by a single psychologist varied from 0 to 12. The a priori chance expectation for number correct is one-half the total number of cadets assigned an A confidence rating in each case. Only two psychologists (L and S) were able to predict the actual adjustment of the cadets better than chance at the 5 per cent level of significance. These results, however, must be interpreted with great caution since seven of the psychologists had *fewer* correct predictions rated A than would be expected by chance alone.

Similarly, when the results obtained for the B and C confidence ratings are considered alone, none of the predictions rises significantly above chance expectation.

Since it seems quite reasonable that greater

[3] One psychologist (R) did make predictions which, when considered alone, were significantly *worse* than chance.

TABLE 3

SUCCESS OF POOLED GLOBAL EVALUATIONS

Condition	Number of Correct Predictions	Chance Expectation of Correct Predictions
Majority agreement	54 out of 100	50
Majority agreement (A ratings only)	17 out of 35	17.5
Unanimous agreement	28 out of 50	25

stability of predictions can be obtained by pooling judgments, the global evaluations made by each psychologist for a given cadet were combined. Two different methods of pooling judgments were used. In one case, whatever was predicted by a simple majority (2 out of 3, or 4 out of 7) of psychologists was accepted as the pooled judgment. A more severe criterion of agreement was obtained by considering only those cadets for whom there was unanimous agreement in Groups 1-4 and a majority agreement of 5 out of 7 psychologists in Group 5. On an a priori basis, the expected number of correct predictions by chance alone is one-half the total number of cadets for whom predictions are available. Inspection of Table 3 clearly reveals that there is little relationship between actual adjustment and predictions obtained by pooling global judgments.

When judgments based upon the single-test approach are examined, the results are essentially the same. Twelve of the clinicians made judgments using the single test approach as well as the global approach. When confidence ratings are disregarded, only 3 of 68 different sets of predictions were sufficiently accurate to prove significant at the 5 per cent level. When only those predictions which the clinician rated A, indicating that he was quite sure of his judgment, are considered, the results are essentially the same. Only one set of predictions, that made by psychologist B using Background Information, is significant at the 5 per cent level. Since a total of 68 significance tests was implicitly made, one for each set of predictions, the apparent success of psychologist B may well be considered a chance fluctuation.[4]

[4] A partial test of this hypothesis was made for psychologist C who did about as well as B on Background Information. C offered to make predictions for two additional groups of 20 cadets each (Groups 2 and

As in the case of the global evaluations, an analysis was made of the results obtained for each separate test, using majority agreement of individual clinicians as a basis for prediction. The number of correct predictions did not differ significantly from chance expectation for any of the six tests. Clearly, there is little relationship between actual adjustment and predictions based upon an analysis of one test at a time.

DISCUSSION

In the present study there is little doubt that the clinical assessments of beginning aviation cadets have no relationship to a criterion of adjustment in the basic flight-training program. The objective findings can hardly be ignored, and it is of utmost importance to consider the possible explanations of these results in order to appreciate fully their significance.

1. *Are the criterion data faulty?* A major factor in the success or failure of any prediction experiment is the validity of the criterion measure. This immediately raises the question "Validity for what?" The present study is concerned with predictions of the adjustment to the stresses of flight training of two groups of aviation cadets carefully selected to reflect highly effective and highly inadequate patterns of behavior. The clinician is called upon to predict (in retrospect) the *future* behavior of each cadet on the basis of test performance obtained when he first entered training. In this sense, the evaluations called for are prognostic of later adjustment rather than diagnostic of present behavior.

The likelihood that our criterion groups of successful and unsuccessful cadets are in error is small. Few, if any, experimental investigations of adjustment to stress have been based upon criteria as objective and manifold as those used in the present study. Although it is true that some members of the successful group may later experience breakdown under high situational stress and may manifest behavior indicative of poor adjustment, the anxiety, functional disorders, and overt personality disturbance which played a prominent role in the failure of each cadet in the unsuccessful group are outright evidence of maladjustment and preclude consideration of these cadets in any other category.

One explanation of the inability of the psychologists to predict the criterion may be the relatively restricted range of adjustment. Since each cadet entering flight training had successfully passed a number of hurdles prior to admission, the likelihood of an individual with severe neurotic or psychotic manifestations being included in the test population is extremely small. This objection has been partially overcome by selecting a small experimental sample of 100 cases from a much larger group in such a way as to reflect only the extremes. This was deliberately done to maximize the likelihood of obtaining significant predictions.

2. *Is it possible that test evaluations were valid with respect to adjustment at the time the test data were gathered, and that later adjustment has little relationship to assessment of initial personality status?* Because of the exhaustive mental, physical, and educational requirements for aviation cadets, an overtly neurotic or disturbed candidate is rare. However, the problem of prediction is to discover latent or subclinical characteristics which constitute predisposition to later manifestation of excessive stress reaction when the actual stresses of military flying are experienced. Hence, if the clinical assessments were concerned with manifest indications of disturbance at the time the tests were administered, it might be expected that they would be unrelated to later adjustment, unless a very disturbed applicant were found. It is possible that this task is unfair to the clinicians before more detailed empirical research data are available on the signs indicative of predisposition. However, at the time the research was undertaken it was hoped that the insights of the clinicians might facilitate the prosecution of the empirical studies. The results of the empirical studies which have been completed, such as those with the Rorschach, sentence completion, and other tests, confirm the conclusion that the items found most discriminating are frequently not those which were expected on a priori grounds.

5), using only cues provided by Background Information. To aid him in improving his subsequent judgments he was given the criterion data for his original set of protocols (Group 1). C first made predictions for the 40 cadets without specific knowledge of his earlier successes and failures. Twenty-five of the 40 predictions later proved to be correct. Then C re-evaluated the 40 test records after examination of the earlier data, and changed several predictions. Twenty-two of the 40 predictions were correct. Since in both cases chance expectation was 20 correct, the null hypothesis could not be rejected.

3. *Were the judges sufficiently familiar with the prediction problem to evaluate the test protocols in a proper context?* The psychologists who participated in this study represented a wide variety of interest and background. Several, while recognized as highly skilled clinicians, had had little or no contact with military populations. The majority, however, had served during the past war as military psychologists, and a few were thoroughly familiar with the problem, having either gone through a flight-training program themselves or worked as psychologists on similar problems for the Air Force. In addition, to ensure some familiarity with the problem, an extensive discussion of the training situation and the types of stress present, together with several representative case reports of cadets who failed to adjust, was forwarded to each participant. The *uniformly* negative results strongly suggest that lack of familiarity with the training situation and the types of adjustment which characterized the criterion groups is not a deciding factor.

4. *Did the judges have sufficient previous experience with the types of test materials to interpret them properly?* Inexperience with the tests used in the Randolph Field battery may have contributed appreciably to the lack of positive findings. To make feasible their administration to large groups, each of the six tests had been adapted from more familiar individual testing procedures. The Ink-Blot Test was a mixture of the commonly accepted individual and group Rorschach tests with one or two minor innovations. The L-D Test was a group form of the Szondi with only one trial rather than several spread over different days as is standard practice when individually administered in a clinical setting. Though similar in structure to other such tests, the SAM Sentence Completion Test (What Is He Saying?) was devised especially for use with student pilots. A large number of the test items, however, were identical to those used in earlier, well-known tests of this type. Except for the lack of an "inquiry," the group form of the Draw-A-Person test was essentially similar to the standard individual procedure. Similarly, the general type of response given in Feeling and Doing and in Background Information was familiar to most clinicians, although the context in which the responses were made was different from the usual clinical setting.

Several participating psychologists have expressed the belief that these modifications of standard procedures in order to make group administration feasible may have changed the meaning of test responses as commonly interpreted in clinical practice. Others have argued that the basic response patterns should remain relatively constant, although less information may be available in the group tests. Whatever the case may be, it is worth noting that many of the psychologists felt they were making meaningful interpretations of the data.

5. *Is it possible that the personality evaluations were largely correct, but the criterion predictions failed due to inadequate integration of the various personality characteristics observed?* The judgment process consists of at least two phases: (*a*) the description of the individual's personality, and (*b*) the weighing of assets and liabilities noted in the individual and the prediction of adjustment to the training program. Since none of the predictions exceeded chance expectations, it is impossible to tell from the predictions alone whether the personality descriptions are invalid or whether they have been used inappropriately in arriving at predictions. In all likelihood both phases of judging have suffered from faulty inferences. Because at least three different psychologists evaluated each cadet, the variation in personality descriptions from one clinician to the next in the kind and degree of personality descriptions written makes a thorough, systematic analysis of this type impossible. Even when only those cases on which there is unanimous agreement as to predicted outcome are considered, there is little congruence of terminology, each clinician favoring concepts related to his own frame of reference. A major problem in further analysis of data of this kind is finding an adequate system for classifying descriptive terms to make comparisons among clinicians.

In spite of such variability, there are underlying consistencies among clinicians which cannot be ascribed to chance alone. The a priori probability that three judges will unanimously agree in their predictions for a single cadet is .25. For the global approach the amount of agreement among the psychologists was significantly better than chance. There was a low but statistically significant agreement in predictions based upon Background Information, Feeling and Doing, and the Drawing Test, when considered individually.

Less reliability was present in predictions based upon the Ink-Blot Test and the SAM Sentence Completion Test. No agreement among clinicians was observed for the L-D Test. Although none of the psychologists was able to predict beyond chance expectation the actual adjustment of the cadets, they did tend to use the test battery as a whole, and to some extent the individual tests, with some degree of consistency.

To see what cues formed the basis for judgments, the evaluations of cadets upon whom there was unanimous agreement were examined. A rating of failure seemed to depend frequently upon the presence of signs of low motivation for flying, immaturity, strong dependency needs, anxiety, and sexual conflict. Ratings of success were less consistent and often based upon conflicting cues. In general, too few clinicians listed cues in sufficient numbers to permit any systematic analysis. Further development and standardization of terms for the description of personality must take place before any appreciable understanding of the process of judging will be possible.

In view of the results of the present study, several questions immediately arise. What are the implications of these findings with respect to the problem of predicting individual adjustment to other life stress situations? Can these results be accepted as reliable evidence of the general invalidity of clinical assessment procedures as presently constituted for coping with prediction problems of this nature? How do the results of this study fit in with the general picture as it is gradually unfolding in somewhat similar studies such as the Michigan VA clinical trainee assessment program?

In the opinion of the writers, the improvement of clinical judgments based on psychological tests and other clinical evidence will depend upon the completion of adequate quantitative, controlled empirical research relating standardized measurement instruments to specific criterion situations. Such situations include diagnostic groups, job criterion groups, and various other life situations. In this respect, the present results are closely related to those of the Michigan program.

REFERENCES

1. ANASTASI, ANNE, & FOLEY, J. P., JR. A research program on the psychiatric selection of flying personnel. V. The Human-Figure Drawing Test as an objective psychiatric screening aid for student pilots. *USAF, Sch. Aviat. Med., Proj. Rep.*, 1952, No. 21-37-002, Rep. No. 5.
2. BARRY, J. R., & RAYNOR, G. H. A research program on the psychiatric selection of flying personnel. Research on the Cornell Index. *USAF, Sch. Aviat. Med., Proj. Rep.*, 1953, No. 21-0202-0007, Rep. No. 2.
3. KELLY, E. L., & FISKE, D. W. *The prediction of performance in clinical psychology.* Ann Arbor: Univer. of Michigan Press, 1951.
4. SELLS, S. B. A research program on the psychiatric selection of flying personnel: I. Methodological introduction and experimental design. *USAF, Sch. Aviat. Med., Proj. Rep.*, 1951, No. 21-37-002, Rep. No. 1.
5. SELLS, S. B., & BARRY, J. R. A research program to develop psychiatric selection of flying personnel. I. Theoretical approach and research design; II. Research progress. *J. aviat. Med.*, 1953, **24**, 29–47.
6. SELLS, S. B., FRESE, F. J., JR., & LANCASTER, W. H. Research on the psychiatric selection of flying personnel: II. Progress report on development of SAM Group Ink-Blot Test. *USAF, Sch. Aviat. Med., Proj. Rep.*, 1952, No. 21-37-002, Rep. No. 2.
7. TRITES, D. K., HOLTZMAN, W. H., TEMPLETON, R. C., & SELLS, S. B. Psychiatric screening of flying personnel: research on the SAM Sentence Completion Test. *USAF, Sch. Aviat. Med., Proj. Rep.*, 1953, No. 21-0202-0007, Rep. No. 3.

Received June 1, 1953.

THE INFLUENCE OF SET ON SEMANTIC GENERALIZATION[1]

NATHANIEL H. EISEN

New York University[2]

SEMANTIC generalization refers to the generalization to other semantically related words of a response originally conditioned to a single word. This is a form of learned, mediated, or secondary generalization, a concept which has become increasingly important to those who attempt to describe complex human behavior in terms derived from Hull's learning theory. For example, Cofer and Foley (1) have used secondary generalization to explain such phenomena as "meaningfulness," set, memory, reasoning, intelligence, and "unconscious anxiety." Dollard and Miller (3) place considerable emphasis on the concept in their discussion of the development of personality and the process of psychotherapy. They attribute the transfer of anxiety to new situations and the development of neurotic symptoms in part to secondary generalization; similarly, the development of the transference relationship during therapy is due in part to "response-mediated . . . mechanisms of generalization." They suggest that a major function of the therapist is to help the patient in identifying his generalized responses and in " . . . attaching the same label or other cue-producing responses to different situations [which] should increase the amount of secondary stimulus generalization and [in this way] help the patient to respond adaptively to culturally emphasized similarities" (3, p. 323).

[1] This article is based on a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at New York University, February, 1952. The writer wishes to express his indebtedness to Professor Leland W. Crafts for his guidance and assistance throughout the course of this study. The writer also wishes to thank Dr. Bernard Locke, of the VA Regional Office, Brooklyn, New York, who provided the psychogalvanometer and the polygraph; Drs. Jacob Cohen and Lawrence Karlin, for advice and assistance with regard to statistical problems; and the many members of the teaching staffs, of the Washington Square College and the School of Education of New York University, through whose efforts subjects were obtained. Above all, the writer wishes to express his gratitude to Benjamin S. Brody, who designed and built the shock apparatus and the various electrodes, who served as assistant during the collecting of the data, and who in many ways made this experiment possible.

[2] Now at the University of Illinois.

The experimental evidence for semantic generalization has been reviewed most recently by Razran (14). A few of the previously reported studies are directly related to the experiment reported here. Razran (11) found that the salivary response conditioned to one word showed greater generalization to words of similar meaning (synonyms) than to words of the same sound (homonyms or homophones). Riess (15), using the GSR as his CR, and a buzzer as his CS, found that adult *S*s showed generalization to antonyms intermediate between that to synonyms and homonyms, thus obtaining a gradient of generalization from synonyms to antonyms to homonyms. Wylie (17), also using the GSR, but shock as her US, obtained maximum generalization to homonyms, with greater generalization to synonyms than to control words. Cofer and Foley (1) explain the discrepancy between Wylie's findings and those of Riess and Razran on the basis of the "intentions," "expectations," and "beliefs" held by Wylie's *S*s as to the purpose of her experiment.

The related concepts of set, attitude, expectancy, and the like have frequently been invoked either explicitly or implicitly to explain experimental findings in this area in a posteriori fashion (e.g., 1, 2, 15, 17). However, few studies have reported the effects of experimental variations in *S*s' set on semantic generalization. Razran (13) found that (*a*) manipulation of "mental sets," induced by practice with lists of controlled associations, could increase or decrease the amount of generalization to words in such semantic relationships to the CS as subordinate, supraordinate, and the like, according to the kind of practice involved; and (*b*) *S*s with different political views differed in the directions to which they showed generalization to sentences reflecting varying political philosophies.

HYPOTHESES

Major hypothesis. The amount of generalization to a word having a given semantic relationship to the CS should be greater when a set has been induced for generalization to that relationship than it is to a word in the same

relationship when a set has been induced for generalization to other relationships.

This hypothesis may be deduced from the Cofer and Foley (1) elaboration of Hullian theory, and it is also consistent with Razran's theoretical position (14). It does not seem to have been stated and tested explicitly, but is implicit in some of the work of Razran (13) and others (e.g., 2). An explicit statement of a related hypothesis is made by Dollard and Miller, although in somewhat different language (3, p. 323).

Minor hypothesis I. Generalization as described above should be significantly greater when an "implicit set" is induced toward a given semantic relationship by having *S* practice in an apparently unrelated task than it is when a set is induced by means of explicit instructions directed toward the same relationship.

Dollard and Miller (3), in their discussion of therapy (pp. 281 ff.), mention three ways in which people may be induced to attach "labels" or other "cue-producing responses" to previously unlabeled situations. These are roughly analogous to three experimental methods for inducing a set: (*a*) The therapist may so arrange matters that the individual spontaneously finds the proper labels himself, which is similar to the condition existing in the experiments by Razran (11) and Riess (15) in which *S*s were permitted to discover the relationships of the generalization stimuli to the CS with no help from *E*. (*b*) He may strengthen some of the patient's random responses by means of "selective repetition" having an "automatic reward effect," which is analogous to the "implicit set" in the experiment reported here. (*c*) He may take an active role in teaching the patient by explicitly pointing out relationships; this is analogous to the explicit instructions used here for inducing a set. According to Dollard and Miller, the first of these methods should be the most effective, because it offers the greatest amount of satisfaction or reinforcement to the patient or *S*, whereas the last method should be the least effective.

Razran (13) has studied the effects on semantic generalization of a set induced by means of practice with lists of controlled associations, but we find no studies in the literature on generalization comparing the effects of different methods of inducing sets.

Minor hypothesis II. Generalization as described above should be greater to words of high frequency value on the Thorndike-Lorge list (16) than to words of low frequency value.

This hypothesis is derivable from Hullian theory, and was indirectly tested by Cofer, Foley, and their associates (e.g., 5) in experiments involving the recall of lists of words. It is also consistent with Razran's work (e.g., 10). However, according to Lashley and Wade, "... the practice required in learning [is] a wholly unreliable measure of the learning process ... So long as the effective attribute is present, the reaction [i.e., generalization] is elicited as an all or none function of that attribute" (9, pp. 74, 81), presumably regardless of the frequency value of the stimuli.

METHOD

Subjects. This experiment required 64 *S*s to meet two criteria for the establishment of a conditioned GSR. In order to obtain this number of *S*s, 95 individuals were seen initially. These were students at various schools of New York University during the spring and summer sessions of 1951. About 50 per cent were volunteers; the remaining were recruited from courses in general psychology in which service as the *S* of an experiment was required. The *S*s were assigned at random to the various experimental conditions.

Of the 64 *S*s finally selected, 40 were males; ages ranged from 17 to 31, with a median age of 23. All had learned English as their mother tongue, or concurrent with it. Questioning at the end revealed that all remained naive as to the purpose of the experiment.

Materials. The conditioned and the generalization stimuli are shown in Table 1. Associated with each CS is a set of four generalization stimuli of high frequency, consisting of a word in each of four semantic relationships (homonyms, synonyms, antonyms, and "neutral" or unrelated control words), and a similar set of low frequency words.

For the preconditioning trial, the words to be used later as CS and as generalization stimuli for a given *S* were interspersed among ten neutral words selected at random. The order of presentation of this list of words during the preconditioning trial was the same for all *S*s.

For the conditioning procedure, 20 high frequency words were selected, each of these having a synonym, a homonym, and an antonym. This group of words, with repetitions of the CS interspersed among them, could be arranged alternately into three different sequences, thus providing three different "problems" for *S*s to "solve" during the conditioning procedure (see below).

For the experimental treatment in which *S*s were given practice in an apparently unrelated task, *E* devised a P (for Practice) test, consisting of 28 words, each with its synonym, antonym, homonym, and associated neutral word arranged in columns in random order.

Apparatus. The psychogalvanometer was manufactured by the Lafayette Instrument Company. It contained a microampere galvanometer with a rated accuracy within 2 per cent. A vernier dial provided scale units of 10 ohms, and a variable shunt permitted

TABLE 1

CONDITIONED AND GENERALIZATION STIMULI AND THEIR "G" COUNT FREQUENCY VALUE*

CS	Homonyms	Synonyms	Antonyms	Neutral
right AA	write AA	correct A	left AA	prepare AA
	rite 10	precise 6	erroneous 4	tuition 2
buy AA	by AA	purchase A	sell AA	country AA
	bye 1	procure 22	dispense 9	robust 4
raise AA	rays A	lift AA	lower AA	promise AA
	raze 2	heighten 5	degrade 8	urban 4
new AA	knew AA	modern AA	old AA	glass AA
	gnu —	unfamiliar 4	obsolete 3	puppet 6

* See reference 16.

the measurable control of the amplification of the GSR changes. The range was 0 to 10M ohms. Connected with the psychogalvanometer was a Lafayette Multichannel Polygraph, model 603, ink recording, with a manual time marker. The GSR electrodes consisted of ½-inch-square zinc plates fitted into ¾-inch-diameter rubber cups containing soft sponge rubber impregnated with electrode paste. Both electrode contacts were on the palm of the same hand. The electrode paste consisted of powdered Bentonite, saturated $CaCl_2$ solution, and glycerine. It was slow drying, showing no change in conductivity in periods of an hour or more.

The stimulus words were typewritten on a thin Cellophane film placed between two sheets of red waxed paper (sold commercially as "Radio-Mat Slides"). Typing was done on a Vari-Type machine in lower case "Copper-Plate Gothic," which provided simple block letters of uniform height. An attempt was made to control for the length of the words by varying the spacing between the letters. The Cellophane sheets were then secured in 35-millimeter Esco Slide Binders. These slides were projected by means of a TDC Projector, model D, 300 watts, fitted with a TDC Selectron Automatic Slide Projector and tray capable of carrying up to 30 slides. A medium gray screen was used, 5½ feet in front of *S*, with the words appearing about 2¼ inches high in an 11-inch × 12-inch field.

Shock electrodes were ¾-inch-square zinc plates attached to an adjustable elastic arm band; these were about 1 inch apart. Contact was reinforced by means of the electrode paste. The shock apparatus drew current from a 120-volt, 60-cycle line, which, by means of a transformer, was changed to 480 volts at a maximum of 40 milliamperes. A resistor and a potentiometer in the circuit permitted variations in the current delivered to *S*, with a maximum of 4 milliamperes.

The *S* sat at a table, with *E*, who operated the projector, at his right. An assistant (*A*) operated the psychogalvanometer and the polygraph. During the conditioning procedure, when the polygraph recorder was not in use, *A* administered the shock and was able to signal *E* as to *S*'s responses without the latter's knowledge.

Experimental design. Simultaneous manipulation of three variables deriving from the three hypotheses called for a 4 × 2 × 2 factorial design, or 16 groups of *S*s. Control of the effects of specific conditioning and generalization stimuli, and of the effects of varying orders of presentation of the generalization stimuli, necessitated rotation of these factors within each group by means of a latin square. The rotation scheme required a minimum of 4 *S*s for each of the 16 groups, or 64 *S*s in all, with no replication.

Procedure. During the brief preliminary period, *S*'s introduction to the GSR apparatus and to shock was accompanied by appropriate but misleading remarks, and his "base resistance level" and response to shock were noted. Following this was a preconditioning trial exposure to the significant stimuli, which were interspersed among a group of neutral words. Here and throughout, each word was projected for 3 seconds, with a 7-second interstimulus interval.

In line with suggestions made by Razran (e.g., 12), the instructions given *S* during the conditioning procedure were designed to motivate him to attend to the stimuli, but to mislead him as to the purpose of the experiment. The CS were presented embedded in word sequences, and *S* was given the task of finding the basis for their arrangement. For example, in Sequence I the CS appeared between words beginning with the same letter ("... dear, ***right***, die, close, wait, ***right***, we..."). Some of the presentations of the CS were reinforced with the US, while *A* kept *E* informed as to whether *S*'s GSR approached the first criterion for conditioning (see below). When necessary, Sequences II and III were introduced, and cues for solving the sequences were given *S*, the aim being to prolong the presentation of the words until conditioning was established, and to terminate this part of the procedure as rapidly thereafter as possible. Later questioning revealed that this goal was achieved.

Reinforcing shock, when administered, was of 1-second duration, and the last one of the three seconds in which the CS appeared. The first 4 or 5 CS presentations were always reinforced, but thereafter reinforcement was partial and random, depending upon the degree to which *S* approached the criterion. The first criterion of conditioning was met when (*a*) a point was reached at which two presentations of the unreinforced CS yielded measurable GSR's, and the preceding and following neutral words showed no response, or (*b*) with a few *S*s for whom the neutral words never failed to elicit a GSR, two presentations of the unreinforced CS gave GSR's at least twice as strong as that of responses to adjacent neutral words.

Half of the *S*s were given the P test as a task ostensibly unrelated either to the conditioning procedure or to the test of generalization to follow. Instructions

were designed to induce a set for synonyms, antonyms, homonyms, or neutral words. The other half of the *S*s worked at a digit-symbol substitution test for 90 seconds, the average time taken by *S*s to complete the P test.

The *S*s who had taken the P test were tested for generalization, with instructions merely to look at the stimuli and to wait for questions about them. The remaining *S*s (those who had not taken the P test) were given explicit instructions to look for words in one of the significant semantic relationships of the CS during the exposure of the generalization series, but to wait for questions before identifying the sought-for word.

After these instructions were given, an additional reinforced CS was presented, followed by the generalization stimuli in prearranged order and, finally, a repetition of the CS without reinforcement. The second criterion for conditioning was met when *S*'s GSR to the final unreinforced CS was larger than that to at least one of the four immediately preceding generalization stimuli.

The *S*s were questioned as to their identification of the generalization stimuli and their expectations of reinforcement. The purpose of the experiment was explained, *S*s' questions were answered, and the need for secrecy was emphasized. The procedure varied in duration from 45 to 65 minutes.

RESULTS

The GSR was recorded as a rising curve on polygraph paper. The height of each curve was measured in $\frac{1}{100}$-inch units and was then transformed mathematically into Logarithm Conductance Change units, as recommended by Haggard (6) and others. It was possible (4) to treat the postconditioning GSR's as a function of the conditioning procedure alone, ignoring the preconditioning-trial responses.

1. Prior to conditioning, there was no significant difference between GSR's to the stimuli to be used as CS and the mean GSR's to the four words to be used as generalization stimuli. Following conditioning, however, the GSR's to the final unreinforced CS were significantly greater ($t = 6.62$, $p < .01$) than the mean of the GSR's to the four generalization stimuli, from which we may infer that conditioning was successfully established.

2. The effects of the attempts to control three variables (CS words used, generalization stimuli used according to their semantic relationships, and the position of the generalization stimuli in their orders of presentation) were investigated by means of three analyses of variance.[3] The only significant F value was that deriving from the interaction of the words used as generalization stimuli, according to their semantic relationships, and the position of these words in their orders of presentation (see Fig. 1). The magnitude of the response to synonyms decreased as a function of the position of the synonym in the order of presentation; that is, the later the synonym appeared in the series of generalization stimuli, the smaller the response to it was apt to be. The other semantic relationships showed no clear-cut trends, and an analysis of the synonym trend indicated that it was not significant. The absence of significance, however, was due to the large within-positions value used as the error term. It should be noted that each of the 16 entries of Figure 1 is the mean of 16 measures, each measure having been obtained from one *S* in each of the 16 experimental groups. A more significant synonym trend might be found if within-position variance were reduced by decreasing the multiplicity of experimental treatments.

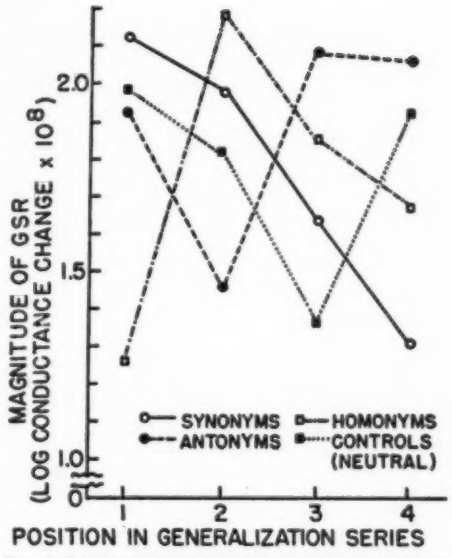


FIG. 1. MAGNITUDE OF THE GENERALIZATION GSR AS A FUNCTION OF THE POSITION OF THE GENERALIZATION STIMULUS IN THE GENERALIZATION SERIES

3. An investigation of the relationship between variations in experimental procedure and *S*'s general response level to stimuli revealed that the groups subjected to different

[3] The supplementary tables have been deposited with the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Order document number 4215 remitting in advance $1.25 for photoprints or $1.25 for 35 mm. microfilm. Make checks payable to Chief, Photoduplication Service, Library of Congress. (See Tables 1, 2; 3, 4; 5, 6.)

treatments were not homogeneous with respect to response level.[4] Whether this difference is attributable to the effects of the experimental treatments themselves or to some bias in the selection of *S*s was not investigated.

4. In the process of testing the various predictions arising from the hypotheses, 36 *t* tests were made. (Four were necessary to test the major hypothesis, 8 each to test the two minor hypotheses, and 16 to test the major hypothesis within each of the 16 experimental groups.[5]) The number of significant, or almost significant, *t*'s was four, just about the number one might expect by chance alone. We therefore cannot reject the null hypothesis that "set" as varied in this experiment did not affect semantic generalization. Alternative statistical treatments, involving nonparametric statistical techniques, yielded similar negative results.

DISCUSSION

We may eliminate the following factors (4) as possible bases for the failure to obtain positive findings, either in the predicted direction or in the form of a generalization gradient: (*a*) lack of sensitivity of the apparatus to intra-individual differences, (*b*) lack of acquaintance of *S*s with the stimuli, (*c*) inadequate conditioning, and (*d*) inadequate statistical treatment of the data. Two factors, however, may explain the obtained results. These are: (*a*) the pattern of reinforcement utilized, and (*b*) the degree of discrimination training achieved during conditioning.

1. It might be argued that owing to differences in the pattern of reinforcements, Riess's (15) *S*s expected shock when exposed to the generalization words, whereas in this experiment the *S*s did not expect shock at this time. Such an explanation seems to call upon some form of "expectancy theory," but is probably translatable into S-R terminology (8).

2. The conditioning procedure was such as to maximize discrimination between the CS and other words. If we assume that complete discrimination did not occur in the other studies (there is no evidence to contradict this assumption), then gradients of generalization are attributable to failure to observe differences among the CS and generalization stimuli. This seems to be consistent with the Lashley-Wade (9) interpretation of generalization. However, the Lashley-Wade position is too inconsistent to serve as a basis for a thorough explanation of generalization. From their discussion it can be argued[6] that the very process of "restriction of associations," which is supposed to bring about generalization, tends to increase discrimination and so tends to decrease the "failure to note distinguishing characteristics" of the stimuli, which is also supposed to increase generalization.

Razran's theory does not account for the results obtained here either. Razran (12, 14) attributes "true generalization" to automatized categorizing, or rating habits, of *S*, habits which increase in strength with "maturity, alertness . . . and other capacity-enhancing influences," and which are correlated, presumably, with the ability to discriminate among stimuli. Although the *S*s here were able, for the most part, to identify the generalization stimuli as related to—and therefore were able to categorize them as similar or dissimilar to—the CS, neither the hypothesized nor any other kind of generalization gradient appeared. Razran introduces the concept of "pseudo-generalization" to account for instances of generalization due to lack of discrimination; but to accept this as an explanatory concept would require the additional assumption that the failure by *S*s of other studies to discriminate among stimuli (and thereby producing a gradient) was the result of fatigue or lack of attention (14).

Hullian theory (7) takes into account the effects of discrimination training on primary generalization. Although the effects of such training on secondary generalization are mentioned by Dollard and Miller (3), the mechanism by which such training operates is not made clear. According to Hull, the degree to which differential reinforcement "steepens" the primary generalization gradient is an inverse function of the proximity of the contrasted stimulus to the CS along the stimulus dimension. By analogy, we should expect minimum effects on the secondary generalization gradient when the contrasted stimulus is an unrelated word (and maximum effects when the contrasted stimulus is a synonym). As an alternative deduction from Hullian theory, it may be stated that secondary generalization

[4] Refer to footnote 3 regarding supplementary tables; see Table 7.

[5] Refer to footnote 3 regarding supplementary tables; see Tables 8, 9, 10.

[6] The writer is indebted to Dr. Lawrence Karlin for suggesting this criticism of the Lashley-Wade position.

may occur along separate dimensions for each of the semantic relationships—a separate gradient for synonyms, antonyms, etc. In that case discrimination training with antonyms should have no effect on generalization to synonyms, and discrimination training with neutral words should have no effect on generalization to other relationships. The conclusion based on either of these two lines of reasoning is the same. Hull's theory does not predict that training with neutral words will result in complete obliteration of the generalization gradient to related words.

CONCLUSIONS

In their preface, Dollard and Miller (3) state that most of their ideas are "hypotheses, the basis for research, not proven principles. Furthermore, it has not been possible to formulate these hypotheses in as rigorous, systematic, and quantitative a form as will eventually be desirable." The results of this study are in agreement with this statement. Although it has been possible to derive hypotheses regarding semantic generalization from the work of Dollard and Miller and others who have been influenced by Hull's thinking, existing theories of secondary generalization cannot account for the results obtained here.

The present findings suggest several new problems and areas of research. Among these are: (*a*) the effects of different degrees of discrimination training on the secondary generalization gradient; (*b*) the effects of discrimination training with stimuli in different semantic relationships of the CS on the gradient of secondary generalization; (*c*) the effects of different *patterns* of partial reinforcement on secondary (and perhaps on primary) generalization; and (*d*) the effects on generalization to the synonym—or any other semantic relationship—of the position in the generalization series at which the crucial word appears. All of these problems are relevant to the use of the concept of semantic, secondary, mediated, or learned generalization as an explanation of what occurs in therapy or in the learning experience of the child.

SUMMARY

The purpose of this study was to compare the influence on semantic generalization of variations in (*a*) the kinds of set given *S*, (*b*) the methods of inducing a set, and (*c*) the past experience of *S* with the generalization stimuli. The findings were essentially negative, generalization occurring randomly and without regard to the relationship of the stimulus to the CS, or of the set given *S*. That is, neither the hypothesized results nor those reported in the literature appeared. Current theories of secondary generalization are inadequate to explain these negative findings, but several new areas of research are suggested.

REFERENCES

1. COFER, C. N., & FOLEY, J. P., JR. Mediated generalization and the interpretation of verbal behavior: I. Prolegomena. *Psychol. Rev.*, 1942, **49**, 513–540.
2. DIVEN, K. Certain determinants in the conditioning of anxiety reactions. *J. Psychol.*, 1937, **3**, 291–308.
3. DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
4. EISEN, N. H. The influence of set on semantic generalization. Unpublished doctor's dissertation, New York Univer., 1952.
5. FOLEY, J. P., JR., & COFER, C. N. Mediated generalization and the interpretation of verbal behavior: II. Experimental study of certain homophone and synonym gradients. *J. exp. Psychol.*, 1943, **32**, 168–175.
6. HAGGARD, E. A. On the application of analysis of variance to GSR data: I. The selection of an appropriate measure. *J. exp. Psychol.*, 1949, **39**, 378–392.
7. HULL, C. L. The problem of primary stimulus generalization. *Psychol. Rev.*, 1947, **54**, 120–134.
8. JENKINS, W. O., & STANLEY, J. C., JR. Partial reinforcement: a review and critique. *Psychol. Bull.*, 1950, **47**, 193–234.
9. LASHLEY, K. S., & WADE, MARJORIE. The Pavlovian theory of generalization. *Psychol. Rev.*, 1946, **53**, 72–87.
10. RAZRAN, G. Salivating and thinking in different languages. *J. Psychol.*, 1936, **1**, 145–151.
11. RAZRAN, G. A quantitative study of meaning by a conditioned salivary technique (semantic conditioning). *Science*, 1939, **90**, 89–90.
12. RAZRAN, G. Attitudinal determinants of conditioning and generalization of conditioning. *J. exp. Psychol.*, 1949, **39**, 820–829.
13. RAZRAN, G. Some psychological factors in the generalization of salivary conditioning to verbal stimuli. *Amer. J. Psychol.*, 1949, **62**, 247–256.
14. RAZRAN, G. Stimulus generalization of conditioned responses. *Psychol. Bull.*, 1949, **46**, 337–367.
15. RIESS, B. F. Genetic changes in semantic conditioning. *J. exp. Psychol.*, 1946, **36**, 143–152.
16. THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Bureau of Publications, Teachers Coll., Columbia Univer., 1944.
17. WYLIE, RUTH C. Generalization of semantic conditioning of the galvanic skin response. Unpublished master's thesis, Univer. of Pittsburgh, 1940.

Received June 1, 1953.

CONDITIONING LEVEL IN THE BEHAVIOR DISORDERS[1]

JANET A. TAYLOR AND KENNETH W. SPENCE
Northwestern University *State University of Iowa*

IN MODERN psychology the clinical concept of anxiety can be considered to be the dominant and key explanatory principle of the behavior disorders. Freud, who is given major credit for originating these theories, stated in 1921 that anxiety was the "fundamental phenomenon and the central problem of the neuroses" (5, p. 85). Although many have disagreed with Freud's notions as to its source and etiology, anxiety as an explanatory concept has been widely accepted (e.g., 2, 3, 7, 19), the current tendency being to regard the behavioral reactions of the neuroses either as direct expressions of anxiety or as a defense against this anxiety. The symptoms found in the functional psychoses have also been similarly analyzed in terms of their relationship to anxiety, although usually not in the detail accorded the neurotic disorders.

The development of the anxiety concept rests primarily on observational evidence gathered from clinical data. While many of the postulated relationships of anxiety to overt neurotic and psychotic behavior appear to have a firm empirical basis, the concept has never been systematically analyzed nor have adequate experimental tests been made concerning many of the speculations about it. It is the purpose of this investigation to attempt to verify experimentally some of the assumptions that have been made.

USAGES OF THE ANXIETY CONCEPT

As the writers have pointed out in previous studies (13, 14) the term "anxiety" appears to be employed by the clinician in at least two different ways: (*a*) as a term applied to a particular set of overt behavior patterns, hereafter referred to as "manifest anxiety," and (*b*) a hypothetical construct, hereafter called "internal anxiety state," that is assumed to be a determinant of overt behavior.

Manifest anxiety. The definition of the concept "manifest anxiety" appears to have been fairly clear and unambiguous. A group of overt reactions, directly observable or measurable, are identified as manifest anxiety. This behavior consists of such verbal responses as reports of chronic worry and tension, and such nonverbal patterns as excessive activity of autonomically controlled effectors, disturbances in appetite and elimination, sleeplessness, panic attacks, etc. These overt behavior patterns are also assumed to be the accompaniment, or to represent the overt expression of, an internal emotional state of the person. While attempts have been made to obtain physiological indices of this state, the emotionality has usually been merely assumed by the clinician to be present, accompanying the more overt, directly observable symptoms mentioned above.

Internal anxiety state. In contrast to this concept of manifest anxiety, which is identified with specific observable behavior symptoms, the concept of internal anxiety is a hypothetical construct that has been introduced in connection with attempts to account for certain other types of behavior patterns that occur in the neuroses. These behavior patterns have typically been referred to as defense reactions, which supposedly protect the individual against the internal state of anxiety. According to this conception the individual learns to respond to emotion-arousing stimulus conditions, not with overt emotional symptoms (manifest anxiety) but instead, with certain instrumental responses. It is assumed, further, that these newly learned responses serve in some manner to avoid the arousal of the internal anxiety state or to reduce its intensity once it has been evoked. Thus the presence of these defense mechanisms is said to result in the minimization of the amount of functional internal anxiety. Since the internal anxiety state is merely *inferred* from these defense reactions, and is not observed or measured, this second usage of

[1] This investigation was supported by a research grant from the National Institute of Mental Health, of the National Institutes of Health, Public Health Service. The writers wish to acknowledge the very helpful cooperation of Dr. David Shakow, Chief of the Department of Psychology, Illinois Neuropsychiatric Institute, and of Dr. Samuel Liebman, Superintendent of the North Shore Health Resort. Acknowledgement is also made to George Faibish and John McAllister, who served as research assistants for the project.

anxiety can be seen to be a hypothetical concept.

Although on a functional level the internal anxiety state is held in check by the presence of the defenses, it should be noted that the individuals exhibiting such symptoms are still considered to be potentially anxious. That is, if the defenses are somehow prevented from occurring, the stimulus conditions previously giving rise to the defense reactions would now arouse the internal anxiety. Furthermore, this anxiety would produce the type of overt behavior that is described as manifest anxiety. This type of interpretation would seem to identify the internal emotional state assumed to be operating at the time of manifest anxiety with the internal anxiety state postulated to lie behind the appearance of defense reactions.

In the typical textbook treatment, the relationship between internal anxiety and overt symptomatology is described for each behavior disorder separately. It can be seen from these discussions that the assumption has not been made by clinical theorists that defenses against internal anxiety operate in uniform fashion and with equal effectiveness in all diagnostic groups, or that the presence of defense reactions precludes the appearance of some manifest anxiety responses. In obsessive-compulsive disorders, for example, manifest anxiety is frequently observed, disappearing only when the obsessive thought and/or compulsive act have occurred. In the classical hysterical syndrome, on the other hand, manifest anxiety is almost completely absent ("la belle indifférence"). Furthermore, the internal anxiety postulated to lie behind the symptoms is kept at a minimal level. Thus, the hysterical defense system may be considered to be particularly effective in checking internal anxiety and the obsessive-compulsive syndrome, in comparison, far less so.

Within the neurotic disorders the symptoms of only one group, anxiety neurotic (anxiety state), have not been assumed to have a defensive function. In fact, this group is often characterized as one in which defense reactions have failed to develop. That is, individuals are classified in this category on the basis of the presence of a large number of manifest anxiety symptoms and the absence of any well-integrated defense system. Thus, while the anxiety neurotic has a few scattered defense reactions, he stands in distinct contrast to individuals in other diagnostic groups, being relatively unprotected against his anxieties.

As we have seen, anxiety plays a key explanatory role in the neuroses. The overt symptoms in these disorders, if not direct expressions of internal anxiety, are defense reactions against it. Thus, the major differences among subgroups are in the particular protective mechanism that is used to reduce the anxiety.

Anxiety has been assigned a similar role in the functional psychoses: many symptoms in these disorders are assumed to have defensive properties with respect to anxiety. However, in contrast to the neuroses, the anxiety concept as an explanatory principle occupies a relatively minor position; anxiety and reactions to it are considered as being only one segment of a multifaceted phenomenon. Furthermore, the theories concerning the dynamics of these disorders are insufficiently detailed to permit direct comparison of neurotic and psychotic groups with respect to amount of internal anxiety. For example, clinical theory would permit us to hypothesize that the anxiety neurotic is more likely to be anxious at any moment of time than the obsessive or the hysteric, but would not provide a basis for similar judgments among psychotic groups or between psychotic and neurotic *S*s.

AN EXPERIMENTAL TEST

A number of studies have appeared in the literature that have shown a relationship between level of classical conditioning and manifest anxiety variously defined, in both normal *S*s (1, 6, 11, 12, 13, 14) and psychiatric patients (9, 17, 18). The interpretation of these results offered by the writers (12, 14) has been that degree of manifest anxiety is related to the intensity of an internal anxiety (emotional) state and that this, in turn, reflects the level of drive (*D*) of a subject. Thus, according to an analysis in S-R terms, *S*s exhibiting much manifest anxiety should, by virtue of higher *D*, exhibit a higher level of conditioning performance than those with lower manifest anxiety. The assumption that manifest anxiety reflects drive level has also been supported by the results of a number of investigations involving learning tasks other than classical conditioning (4, 8, 16).

Attention was directed to the fact that clini-

cians have tended to postulate that the internal anxiety that produces manifest anxiety symptoms is the same state as that responsible for evoking defensive reactions. If we follow this assumption, we see that the established relation between manifest anxiety and conditioning provides a basis for testing the amount of functional internal anxiety that is present in patients who do not exhibit manifest anxiety. That is, differences among subgroups of patients in level of conditioning could be interpreted as reflecting differences in degree of functional internal anxiety.

In an earlier study by the writers (13) involving comparisons between normal, neurotic, and psychotic *Ss*, conditioning scores were used in this manner. Clinical speculations led to the prediction that psychotic and neurotic *Ss* would, in general, exhibit a higher conditioning level than normals due to the greater degree of manifest and internal anxiety in these individuals. The results of this investigation indicated that the conditioning performance of the psychotic *Ss* was significantly higher than that of both normals and neurotics. The neurotic group was superior to the normal *Ss* but the difference was not statistically significant.

The purpose of the present experiment was to test further implications of clinical theory. Specifically the prediction was made that *Ss* with the diagnostic label "anxiety state" should show higher conditioning scores than *Ss* in other neurotic groups in which defense behavior is said to predominate.

METHOD

Subjects. The *Ss* were psychiatric patients obtained from the inpatient and outpatient services of the Illinois Neuropsychiatric Institute and from the North Shore Health Resort. Patients were used as *Ss* in the experiment as they became available without any prior selection or screening as to diagnosis. Some screening was involved, however, in that the psychiatrists referred only those patients whom they judged to be capable of cooperating in the experimental situation and who would not be excessively disturbed by the experience.

Psychiatric diagnoses for the *Ss* were drawn from the individual files maintained by each institution. In addition to the diagnosis, other types of information were obtained for each patient. These included items of personal data (age, occupation, onset of illness, etc.) drawn from case histories and a psychiatric check list that was completed for each patient by the psychiatrist in charge of the case. The check list consisted of statements describing common behavior patterns found in the functional behavior disorders, including a detailed list of overt anxiety symptoms. The psychiatrist was asked to indicate absence or severity of each symptom if present, and further, to assign to each patient an over-all rating as to the degree of overt anxiety. In addition, an individual form of a manifest anxiety scale developed by one of the writers (15) was administered to each patient, usually several days after the conditioning session.

TABLE 1

DISTRIBUTION OF DIAGNOSES IN THE NEUROTIC AND PSYCHOTIC GROUPS

Neurotics		Psychotics	
Diagnosis	*N*	Diagnosis	*N*
Anxiety state	15	Schizophrenia	
Anxiety hysteria	5	Simple	4
Obsessive-compulsive	3	Paranoid	12
Depressive reaction	10	Hebephrenic	3
Hysteria (misc.)	5	Mixed type	12
Hypochondriasis	2	Involutional psychosis	3
Psychoneurosis mixed	17	Miscellaneous	8
Alcoholism	7		
Drug addiction	4		
Behavior & character disorders	6		
Total	74	Total	42

One hundred and forty-five patients participated in the conditioning situation. The records of 29 of these were excluded from consideration in the results of the investigation for the following reasons: 6 for equipment failure, 3 due to inability or unwillingness to cooperate in the conditioning situation, 2 because of unavailability of any diagnosis, and 18 because they did not meet the latency criteria set up to exclude voluntary responders. Of the 116 remaining *Ss*, 74 were diagnosed as neurotic and 42 as psychotic.[2] A more detailed breakdown of the composition of the neurotic and psychotic groups by diagnostic category is shown in Table 1.

Apparatus and method of recording. The equipment for recording the eyelid closure and presenting the CS and UCS has previously been reported (13) and will not be described here. Seated in a straight-backed chair in a semi-darkened room containing the conditioning apparatus, *S* was instructed to blink upon receiving a verbal ready signal and then to fixate on a 6-cm. circular milk glass window or disc, placed at a distance 6 ft. in front of him. The brightness of the disc between trials was approximately 1 apparent foot-candle. The CS consisted in an increase in brightness of the disc to approximately 24 apparent foot-candles. The duration of the CS on each trial was 520 msec. Following the onset of the CS by 470 msec., the UCS, a puff of air, was administered to the right eye. The air puff was produced by the fall of a 30-mm. column of mercury, magnetically released, in a manometer. Intertrial inter-

[2] Of this total, the data from 34 neurotic and 21 psychotic *Ss* were presented in an earlier study (13) by the writers. Additional *Ss* were tested and the entire group utilized in the analyses reported in the present investigation.

vals of 15, 20, and 25 sec., averaging 20 sec. and fixed according to a prearranged schedule, were used.

Studies in the Iowa laboratory have shown that some *Ss* attempt to avoid the air puff by voluntarily closing their eye to the CS, keeping it closed until after the occurrence of the UCS. In order to eliminate such voluntary responders, objective latency criteria utilized in previous studies (10, 13) were employed. Briefly, the records of all *Ss* who gave more than 50 per cent CR's with latencies less than 300 msec. were excluded from the sample.

Conditioning procedure. Each *S* first received three presentations of the light alone in order to check for any initial conditioned response tendencies to the light. A single puff of air, unaccompanied by the light, was then administered and spontaneous blinking recorded for 40 sec. Following these preliminary trials, all *Ss* were given 80 paired presentations of the light and air puff.

RESULTS AND DISCUSSION

Comparison of neurotic groups. In comparing the conditioning scores of anxiety neurotics with those in other neurotic groups a question arose as to whether or not certain diagnostic categories should be included in the analysis, namely: behavior disorders, alcoholism, drug addiction, and mixed psychoneurosis. The latter diagnosis is typically a "wastebasket" category into which individuals failing to correspond to any of the other neurotic patterns are placed. Since many of these *Ss* may be expected to exhibit a number of overt anxiety symptoms in addition to other reactions, opposing this group to anxiety neurotics might not be considered a fair test of the clinical theory. The other three groups (alcoholics, behavior disorders, and drug addicts) likewise raised a problem since the relationship of their behavior to anxiety has not been clearly specified.

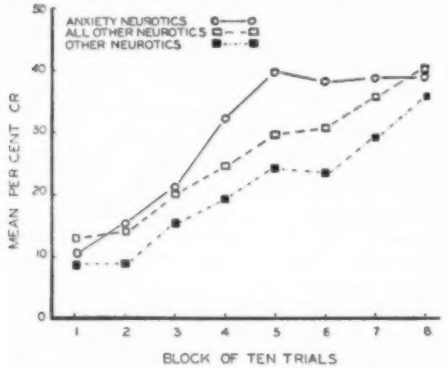


FIG. 1. ACQUISITION CURVES SHOWING THE MEAN NUMBER OF CR'S FOR THE NEUROTIC GROUPS IN SUCCESSIVE BLOCKS OF TEN TRIALS

Accordingly, two analyses were made between the conditioning scores of anxiety neurotics and those of other groups, one including all other neurotics and the other with the four questionable categories omitted.

Curves of acquisition of the CR for anxiety neurotics and the other groups are shown in Fig. 1. Although the curve for anxiety neurotics tends to be above that for all other neurotics, as would be predicted by clinical theory, there was some overlapping. The mean conditioning scores for anxiety neurotics were also slightly higher than for all other neurotics, as is shown in Table 2. The difference between means, however, was not statistically significant.

When the scores for the individuals in the four categories previously mentioned were excluded from the nonanxiety category, the conditioning scores for this group were decreased. Although these *Ss* ("other neurotics") exhibited a lower conditioning performance on every block of trials than did anxiety neurotics, the t ratio comparing the mean conditioning of this reduced group with anxiety neurotics was only 1.22, a value falling far short of the level needed to reject the null hypothesis.

In attempting to explain the failure to find significant differences between the neurotic groups a check was made on the fundamental assumption that the presence of the defense reactions precludes, or at least reduces, the appearance of manifest anxiety symptoms. On a questionnaire filled out for each patient by his psychiatrist the patient was classed into one of three groups according to the over-all amount of manifest anxiety exhibited: absent or mild, medium, or severe. (All judgments were made in terms of a patient population.) A chi-square test of independence indicated that there was no significant difference between the number of anxiety neurotics and all other neurotics assigned to each of these categories. When the four questionable groups were removed, the anxiety ratings for the latter group decreased slightly but the value of chi square again failed to reach statistical significance. A comparison of means on the manifest anxiety scale, described earlier, also failed to show differences, the mean of the anxiety neurotics being 30.00 and of all other neurotics 30.96. Since the groups were similar in these measures of overt anxiety, it is not surprising that their conditioning performance also did not differ significantly.

The results of the present investigation, then, do not conform to the predictions developed from the clinical hypotheses concerning the significance of certain neurotic symptoms. Not only was there failure to support the inference that these symptoms reduce internal anxiety, as shown by the conditioning data, but also the analyses of the estimates of manifest anxiety indicated that the groups did not differ in this respect. In this group of patients at least defensive neurotic symptoms seemed to be merely added to manifest anxiety reactions and did not replace them.

A number of qualifications could be added to the defense theory which would leave it essentially intact and still explain the present findings. Individual differences might be expected in the degree to which a defense system was perfected. For example, patients whose symptoms have appeared only recently might still be in the process of developing them and hence be more anxious than persons who had a long history of illness; some individuals might be attempting to give up their neurotic symptoms and others might be content with them. Failure to take such variables into account might obscure the relationship of well-developed defenses to anxiety. Or it might be claimed that in terms of the number and efficiency of the defenses, neurotic disorders lie along a continuum in such a fashion that opposing one extreme group with all others, as was done in the present study, will reveal no differences in either manifest or internal anxiety unless extremely sensitive measuring instruments are used. Whatever the explanation, a reasonable conclusion to be drawn from the results of the present study is that the clinical hypotheses describing the significance of neurotic behavior are, at best, oversimplified and need elaboration. Further research, testing such speculations as those offered above, is certainly indicated.

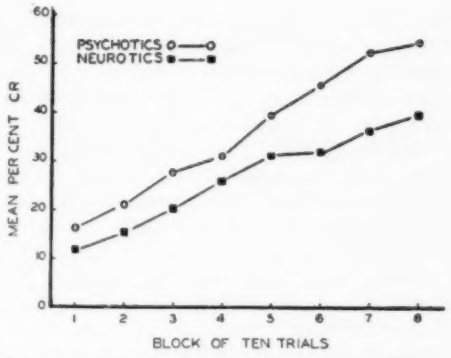


FIG. 2. ACQUISITION CURVES SHOWING THE MEAN NUMBER OF CR'S FOR THE NEUROTIC AND PSYCHOTIC GROUPS IN SUCCESSIVE BLOCKS OF TEN TRIALS

Psychotics vs. neurotics. The *S*s in the present investigation include 34 neurotic and 21 psychotic individuals whose conditioning performance was reported by the writers in another study (13), plus an additional number of *S*s tested at a later time. In the original study the psychotic *S*s, when compared to neurotics, exhibited a significantly superior conditioning performance. In order to determine whether this difference would still be found with a larger sample from each population, conditioning scores for the groups were again compared. The conditioning curves shown in Fig. 2 indicate that the psychotic group continued to exhibit a consistently higher performance throughout the course of the conditioning trials. However, as may be seen from Table 3, the results of a *t* test indicated that the difference between means for the total conditioning scores was significant only between the .05 and .10 level of confidence, somewhat lower than was found in the earlier study.

The two groups were also compared with respect to the data from the psychiatric questionnaire and the manifest anxiety scale. For

TABLE 2

MEAN PERCENTAGE OF CONDITIONED RESPONSES EXHIBITED BY NEUROTIC GROUPS IN 80 CONDITIONING TRIALS

Group	*N*	Mean	σ_M
Anxiety neurotics	15	28.67	5.98
All other neurotics	59	26.18	3.32
Other neurotics minus four categories	25	20.50	2.90

TABLE 3

MEAN PERCENTAGE OF CONDITIONED RESPONSES ON 80 CONDITIONING TRIALS

Group	*N*	Mean	σ_M	*t*
Neurotics	74	26.69	2.98	1.82*
Psychotics	42	35.89	4.11	

* Significant between the .05 and .10 level of confidence for 104 *df*.

the former, a chi-square test of independence indicated no difference in the proportion of neurotic and psychotic *Ss* in each of the three overt anxiety categories. A comparison of the means on the manifest anxiety scale also revealed no significant difference between groups. Thus, it would appear that if the superiority of the psychotic group in conditioning were to be interpreted in terms of anxiety, the differences would be in the amount of functioning internal, rather than manifest, anxiety. According to this interpretation, psychotic symptoms mask or partially prevent the appearance of overt anxiety responses without a corresponding reduction in the amount of internal anxiety.

A further possibility that should be mentioned is that the greater conditioning performance of psychotics may be due to factors other than anxiety. While previous studies have demonstrated a relationship between conditioning and manifest anxiety, variously defined, the correlation coefficients that have been reported between these two variables (6, 13) indicate that a relatively small amount of the variability among *Ss* can be accounted for in terms of anxiety. Thus, it is logically possible that the superiority of psychotics over neurotics, and, as was found in an earlier study (13) over normals, is wholly or partially due to as yet unidentified factors that separate psychotic individuals from other groups.

SUMMARY

The present study was designed to investigate differences in conditioning performance between anxiety neurotics (anxiety state) and neurotics in other diagnostic categories, and between neurotics and psychotics. Seventy-four neurotic *Ss* (including 15 anxiety neurotics) and 42 psychotic *Ss* were run in an eyelid conditioning situation under the same experimental conditions.

The conditioning performance of the anxiety neurotics and other neurotics did not differ significantly. Further, no difference between the two groups was found in the amount of overt anxiety exhibited, as measured by a test of anxiety and a psychiatric questionnaire filled out for each patient. The implications of these results for clinical hypotheses concerning the function of certain neurotic symptoms were discussed.

When the psychotic *Ss* were compared to the total neurotic group, their conditioning performance was found to be superior between the .05 and .10 level of confidence. Speculations concerning this difference were offered.

REFERENCES

1. BITTERMAN, M. E., & HOLTZMAN, W. H. Conditioning and extinction of the galvanic skin response as a function of anxiety. *J. abnorm. soc. Psychol.*, 1952, **47**, 615–623.
2. BROWN, J. F. *Psychodynamics of abnormal behavior*. New York: McGraw-Hill, 1940.
3. CAMERON, N. *The psychology of behavior disorders: a biosocial interpretation*. Boston: Houghton Mifflin, 1947.
4. FARBER, I. E., & SPENCE, K. W. Complex learning and conditioning as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 120–124.
5. FREUD, S. *The problem of anxiety*. New York: Norton, 1936.
6. HILGARD, E. R., JONES, L. V., & KAPLAN, S. J. Conditioned discrimination as related to anxiety. *J. exp. Psychol.*, 1951, **42**, 94–100.
7. MAY, R. *The meaning of anxiety*. New York: Ronald, 1949.
8. MONTAGUE, E. K. The role of anxiety in serial rote learning. *J. exp. Psychol.*, 1953, **45**, 91–96.
9. SCHIFF, E., DOUGAN, C., & WELCH, L. The conditioned PGR and the EEG as indicators of anxiety. *J. abnorm. soc. Psychol.*, 1949, **44**, 549–552.
10. SPENCE, K. W. Learning and performance in eyelid conditioning as a function of the intensity of the UCS. *J. exp. Psychol.*, 1953, **45**, 57–63.
11. SPENCE, K. W., & FARBER, I. E. Conditioning and extinction as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 116–119.
12. SPENCE, K. W., & TAYLOR, JANET A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.
13. SPENCE, K. W., & TAYLOR, JANET A. The relation of conditioned response strength to anxiety in normal, neurotic and psychotic subjects. *J. exp. Psychol.*, 1953, **45**, 265–272.
14. TAYLOR, JANET A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–89.
15. TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.
16. TAYLOR, JANET A., & SPENCE, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44**, 61–64.
17. WELCH, L., & KUBIS, J. Conditioned PGR (psychogalvanic response) in states of pathological anxiety. *J. nerv. ment. Dis.*, 1947, **105**, 372–381.
18. WELCH, L., & KUBIS, J. The effect of anxiety on the conditioning rate and stability of the PGR. *J. Psychol.*, 1947, **23**, 83–91.
19. WHITE, R. W. *The abnormal personality*. New York: Ronald, 1948.

Received June 5, 1953.

THE ROLE OF EXPECTANCY IN THE PERFORMANCE OF POSTHYPNOTIC BEHAVIOR[1]

SEYMOUR FISHER

Army Medical Service Graduate School

DESPITE a wealth of anecdotal material and case reports, there have been few experimental investigations of the performance of posthypnotic behavior. Aside from an important contribution by Erickson and Erickson (3) and several studies dealing with the duration of posthypnotic suggestions summarized in Weitzenhoffer's paper (8), the major variables remain essentially unexplored.

The purpose of the present communication is to report the results and implications of a simple, but relatively well-controlled, experiment on posthypnotic suggestion. The primary variable under investigation was the nature of the setting in which the suggestion was carried out, and an attempt was made to manipulate the subject's (*S*'s) expectancies *after* the termination of the formal hypnotic state. Thus, the general hypothesis being tested is that successful performance of a posthypnotic act is not only a function of the specific suggestions given *under* hypnosis, but is also contingent upon the *S*'s subsequent expectations.

METHOD

Subjects

A total of 13 *S*s was employed in this study. All were capable of attaining either a medium or deep hypnotic state, and all had previously shown that they could respond to simple posthypnotic suggestions. Twelve *S*s were males who had earlier participated in a separate hypnotic experiment; the thirteenth *S* was a high school girl who had had only a few hours of hypnotic experience prior to this experiment.

Procedure

While under hypnosis each *S* was given the following posthypnotic suggestion, which was generally repeated two or three times:

"Now listen carefully. After I awaken you, you'll notice a very interesting thing occurring. Every time you hear the word 'psychology' you'll reach up and scratch your right ear. You will be wide awake, and aware of everything that's going on, but each time the word 'psychology' is mentioned, you'll feel compelled to scratch your right ear. Do you understand that clearly?"

During the administration of the suggestion, *S* was alone in the room with the hypnotist (hereafter referred to as E_1). At the conclusion of the suggestion, *S* was awakened and the experiment proceeded as follows:

Period 1. One or more experimenters (associates of the hypnotist who were collaborating in the study) entered the room and were introduced to *S*. A general discussion ensued which included *S*, E_1, and whoever else was present. Each of the various *E*s managed to weave into the conversation the word "psychology" a number of times (a minimum of five), and also attempted, by means of indirect questioning, to get *S* himself to say the word. While the discussion was carried on in a friendly, informal, nonexperimental atmosphere, it was apparent at all times that the *E*s were taking note of *S*'s ear-scratching behavior.

Period 2. After the discussion had gone on about ten minutes, an attempt was made to alter inconspicuously the nature of the situation. The other *E*s temporarily dropped out of the conversation, and E_1 began to chat with *S* about the posthypnotic suggestion which had been given him under hypnosis. The ancillary *E*s soon joined in this discussion, and during the next five or ten minutes every effort was made to create and sustain the impression that the "experiment" was completed—i.e., that the *E*s had satisfactorily witnessed the effects of the posthypnotic suggestion, and were now no longer interested in ear-scratching behavior per se. Throughout this period, however, the stimulus word "psychology" was again mentioned by all the *E*s on numerous occasions.

In order to clarify the technique used during Period 2, the following is a truncated account of the conversation with one *S*:

E_1: "Have you noticed anything unusual during the past few minutes, Jim?"

S: "Why, yes, I've been scratching my ear every once in a while."

E_1: "Uh-huh, as you may have guessed, you were responding to a posthypnotic suggestion I gave you earlier. Did you notice *when* you were scratching your ear?"

S: "It seems like it was to a word or something."

E_1: "That's right. I had given you the suggestion that every time you heard the word 'psychology' you would scratch your ear. How did you feel the first time you heard the word 'psychology'?"

S: "Well, I just found my hand reaching up."

E_1: "Yes, posthypnotic suggestions work that way. All that's necessary is to tell you while you're in hypnosis that you'll scratch your ear whenever you hear the word 'psychology,' and sure enough you find yourself doing it."

The conversation then continued along the same lines with the other *E*s joining in.

Period 3. As an additional control, E_1 now endeavored to restructure *S*'s expectancies by intimating that the experiment was still in progress. The hypnotist

[1] Based on one portion of a dissertation submitted to the faculty in partial fulfillment of the requirements for the Ph.D. degree at the University of North Carolina, 1952.

TABLE 1

PRESENCE (+) OR ABSENCE (−) OF A POSTHYPNOTIC RESPONSE OF EAR SCRATCHING TO HYPNOTIST'S PRESENTATION OF STIMULUS WORD

Subject	Period 1	Period 2	Period 3
Co	+	−	+
Pa	+	−	+
Ho	+	−	−
Ro	+	−	+
He	+	*	+
Mo	+	−	+
Ea	+	−	+
Bl	+	+	+
Bu	+	+	+
Wa	+	−	+
Ed	+	−	+
Me	+	†	+
Ha	+	−	−

* This *S* failed to respond the first three times the stimulus word was presented, and then suddenly began to respond.

† This *S* held her hand against her ear throughout the second period.

informed *S* that he was going to rehypnotize him and remove the suggestion:

"All right, Jim, in a few moments I'm going to hypnotize you once again and remove the suggestion I gave you. I'm sure you wouldn't want to go around scratching your ear indefinitely. Before I put you back in hypnosis, though, I was just wondering . . . do you remember the name of the last psychology book you bought?"

The other *E*s in the room then presented the stimulus word a few times (both in sentence form and as a single, isolated word), and finally *S* was directly requested to say the word himself. Regardless of whether or not he responded to the stimulus, *S* was subsequently rehypnotized and the suggestion was permanently removed.

At a later date, ranging from one day to a week, those *S*s who failed to respond during Period 2 were interviewed individually by E_1 alone. In this interview interest was focused upon *S*'s attempts to account for the differences in behavior between Periods 1 and 2.

RESULTS

Table 1 records the presence (+) or absence (−) of the ear-scratching behavior in response to the stimulus word "psychology" as presented by E_1 during the three periods of the experiment. As was readily expected, all thirteen *S*s responded positively during Period 1.[2] Under the altered conditions of Period 2, however, nine *S*s completely stopped responding; two *S*s (Bu and Bl) continued to respond in the same manner as they did in Period 1; and one *S* (Me) confused the issue considerably by keeping her hand against her ear throughout the second period, making it difficult to determine whether or not she was responding. One *S* (He) failed to respond in Period 2 the first three times the stimulus word was mentioned by E_1, but then abruptly resumed the response and continued for the remainder of the period.

Table 1 further indicates that, in Period 3, seven of the nine *S*s who ceased responding in Period 2 reverted to the ear-scratching behavior in the absence of any intervening formal hypnosis.

Of particular interest are the data pertaining to the variability in response consistency. That the instructions were interpreted individualistically by different *S*s is clearly seen in an examination of Table 2. While all the *S*s without exception responded in Period 1 when the stimulus word originated from E_1, four *S*s did not respond when they themselves uttered the word; two of these four, moreover, failed to respond when any of the ancillary *E*s presented the cue. Table 2 shows that five *S*s could not be induced to say the word "psychology," and it is noteworthy that all *S*s manifested a decided reluctance to say the word.

One of the most striking findings was ob-

TABLE 2

PRESENCE (+) OR ABSENCE (−) OF A POSTHYPNOTIC RESPONSE OF EAR SCRATCHING RELATIVE TO THE SOURCE OF STIMULATION DURING PERIOD 1

Subject	Source of Stimulation		
	E_1	Other Experimenters	Self*
Co	+	+	−
Pa	+	+	?
Ho	+	−	−
Ro	+	+	?
He	+	+	+
Mo	+	+	+
Ea	+	+	−
Bl	+	+	?
Bu	+	−	−
Wa	+	+	+
Ed	+	+	+
Me	+	+	?
Ha	+	+	?

* A question mark (?) indicates that *S* could not be induced to say the stimulus word.

[2] Although most *S*s were completely consistent in either responding or not responding during any given period, two *S*s during Period 1 missed the cue on single occasions.

tained in the postexperimental interview when the *S*s were confronted with the fact that they had stopped scratching their ear during Period 2. Of the nine *S*s interviewed, two (Co and Ro) offered elaborate "rationalizations": one said he noticed his ear itching severely each time the word was mentioned, but could not scratch it because his hands were in his pockets (which was untrue); the other claimed he was unable to respond because he had a book in his hands at the time (equally untrue). The remaining seven *S*s insisted they did not hear the word mentioned during Period 2!

DISCUSSION

Nature of posthypnotic behavior. A comparison of the behavior in the three experimental periods strongly suggests that the performance of a posthypnotic suggestion is a function of *S*'s understanding of the hypnotist's expectations—not only when the suggestion is given during the formal hypnotic state, but also in the process of enacting the suggestion. In Period 1 *S*s responded each time the signal was perceived, and it may be safely assumed that this behavior would have continued if the hypnotist had not restructured the situation in Period 2. Although *S* was allegedly "wide awake," obviously some aspect of the hypnotist's behavior affected *S*'s posthypnotic behavior. To the extent that the hypnotist was successful in manipulating the experimental periods, it seems reasonable to assume that *S*s who failed to respond in Period 2 were led to believe that the ear-scratching behavior was no longer expected. While the actual suggestion stipulated nothing about the termination of the behavior, and specifically directed *S* to respond *every* time the word was mentioned, the majority of the *S*s were, so to speak, "caught off guard." When the hypnotist implied at the beginning of Period 3, however, that the suggestion was supposed to be still operative, *S*s tended to revert to the posthypnotic act.[3]

[3] Further support for the validity of this interpretation is found in a hypnotic interview held with one *S* (He) at the conclusion of the experiment. When asked, under hypnosis, to account for the fact that in Period 2 he failed to respond the first three times the word was presented but then abruptly began to respond again, after extensive questioning he stated that he suddenly realized that "the experiment wasn't over yet." One might hypothesize that had it been possible to structure the situation even more precisely during Period 2, all 13 *S*s might have been "caught off guard."

The fact that a posthypnotic suggestion is carried out only in a specific context has never been sufficiently recognized in the literature. Indeed, it has been generally implied that, however unexpectedly a posthypnotic signal may arise, the good *S* finds himself responding (1, 4, 6).

A number of studies indicate that when an *S* is given a posthypnotic suggestion without a terminal cue (i.e., indicating when *S* should stop responding), the behavior tends eventually to cease as a function of time (e.g., 5, 7). That this behavior, however, can be "turned off and on" by controlling *S*'s expectations in the subsequent "waking" state seems to be a hitherto unrecognized phenomenon which refocuses the entire problem of posthypnotic suggestion. In effect, the question may be posed: Is *S* in some form of hypnotic state during the *total* time that he is prepared to respond posthypnotically, even when the behavior extends over a period of days without any intervening formal hypnosis? The answer to this question is certainly not a simple one, but will be touched on very briefly here.

Erickson and Erickson (3) have elaborated on Moll's earlier observation (6) that *S* frequently appears to pass into a spontaneous hypnotic state immediately preceding the suggested posthypnotic act; this spontaneous state terminates as soon as the behavior is completed, and *S* resumes his "normal" activity. The specific manifestations of the state are thoroughly described by the Ericksons, who also point out that this new state of hypnosis can be most clearly demonstrated by propitiously interrupting *S* in the process of performing the posthypnotic act: *S* will now heed the hypnotist's suggestions once again, although both before and after the response it is claimed that *S* is no more suggestible to the hypnotist's requests than is a nonhypnotic *S*.

In the present study the manifestations of the spontaneous, self-limited state were clearly observed[4]—but only on each occasion when *S* began to reach up to scratch his ear. That is, during the greater part of the interaction between *S* and *E*s, there were no indications of

[4] Amnesia for the ear-scratching behavior, however, did not occur even in those *S*s who were ordinarily amnesic for hypnotic events. It is probable that the discussion during Period 2, in which specific reference was made to the ear-scratching, contributed to the lack of subsequent amnesia.

any hypnotic-like state; whenever the stimulus word was effective, however, the new hypnotic state became momentarily evident until *S* finished scratching his ear. Now if it is assumed that *S*s were "wide awake and normal" in between the ear-scratching responses, then it is difficult to see why the behavior changed in Periods 2 and 3. When the hypnotist restructured the setting at the beginning of Period 2, there were no indications of a hypnotic state at that moment; and certainly when *S*s resumed the ear-scratching behavior in Period 3, no formal hypnosis was instigated.

The results of this study suggest, rather, the view that during the *entire* time *S* remained in the setting after being told to awaken, a certain degree of rapport existed between him and the hypnotist. This rapport was quite specific, it appears, in the sense that *S* continued to be concerned about the hypnotist's expectations regarding the posthypnotic behavior previously suggested. In no other way did he appear hypersuggestible.

This concern over the hypnotist's expectations is further seen in the denial behavior observed during the postexperimental interview with the hypnotist. Although the tendency for hypnotic *S*s to rationalize a posthypnotic suggestion is commonplace, it appears that a hypnotic *S* is equally prone to rationalize his *lack* of posthypnotic behavior—particularly if he believes he has failed to live up to the hypnotist's expectations.[5] Moreover, when one considers the fact that these denials and rationalizations were obtained from some *S*s as long as one week after the last formal hypnotic session, it becomes apparent that the line of demarcation between termination of hypnosis and "normal waking behavior" is a hazy one indeed. Some form of rapport seems to persist between *S* and the hypnotist since the defense mechanisms manifested by *S*s reflect a degree of ego involvement which would not be expected in the absence of a strong (hypnotic?) relationship. One rather obvious question comes to mind: Does *S ever* completely resume his prehypnotic, "normal" state, where he is no longer affected by the hypnotist? A promising area of research might deal with the type and duration of the subject-hypnotist relationship after conclusion of formal hypnotic sessions.

Interpretation as a function of expectancy. In a paper by Erickson (2) it is claimed that the hypnotic *S* interprets the hypnotist's suggestions in a concrete, literal manner. The present data are clearly not in line with her contention, since *S*s do not attach any literal meaning to the phrase "every time the word 'psychology' is mentioned." Instead, it would seem that whatever interpretation is given to the explicit suggestion depends upon what *S* believes the hypnotist intended. It will be recalled that when the suggestion was presented during hypnosis, *S*s did not know that additional people were to be later introduced. In Period 1, after *S* is awakened, it is almost as though he has to decide "does he (the hypnotist) expect me to respond to these people also?" The fact that eleven out of the thirteen *S*s did respond to the ancillary *E*s may possibly be explained by noting that these *E*s were presented to *S*s by the hypnotist and might thus be considered as "associates." The prediction could be made that if the hypnotist ignored, or in some way appeared antagonistic toward, the other *E*s, fewer *S*s would respond to them as a source of stimulation.

Similarly, it seems doubtful whether it occurred to *S*s during the formal hypnosis (when they were given the posthypnotic suggestion) that they themselves might be asked to present the stimulus word. The peculiar difficulty in getting *S*s to say the word "psychology" during Period 1 cannot be readily accounted for by the present data, but is consistent with the possibility that *S*s were so uncertain as to the hypnotist's expectations on this score that they were reluctant to have to commit themselves one way or the other. This phenomenon is an interesting one, and warrants further research.

SUMMARY AND CONCLUSIONS

The primary purpose of this study was to investigate the extent to which the effective enactment of a posthypnotic suggestion is dependent upon the nature of the setting in which the suggestion is carried out. On the basis of data obtained from thirteen hypnotic *S*s, the results were as follows: (*a*) The continued performance of a posthypnotic suggestion is a function of *S*'s belief that the hypnotist expects the behavior to occur; i.e., if *S* is "caught off guard," the posthypnotic behavior

[5] By implication, it follows that introspective reports of *S*s who have been used in hypnotic experimentation may be of dubious validity and should be interpreted with extreme caution.

tends to disappear. However, additional manipulation of the situation by the hypnotist can reinstate the behavior in the absence of any further formal hypnotic suggestion. (*b*) The hypnotic *S* does not usually tend to interpret the hypnotist's suggestions in a concrete, literal manner, but rather, the eventual meaning appears to be determined by *S*'s inferences as to the hypnotist's expectations. (*c*) An incidental finding of some interest indicates that *S*s who are "tricked" into a lapse of carrying out a posthypnotic suggestion seem to find it necessary to "save face," either by a complete denial of their "negligence" or by a more elaborate form of rationalization. The tendency for hypnotic *S*s to rationalize their posthypnotic behavior is well-established in the literature; it appears, however, that hypnotic *S*s are equally capable of rationalizing their *lack* of posthypnotic behavior.

The results suggest that when an *S* is carrying out a repetitive posthypnotic suggestion over an extended period of time, a certain degree of rapport continues to exist between him and the hypnotist. This rapport is evidenced by the *S*'s unceasing concern over the hypnotist's expectations in regard to the previously suggested posthypnotic behavior.

REFERENCES

1. BERNHEIM, H. *Suggestive therapeutics.* New York: London Book Co., 1947.
2. ERICKSON, ELIZABETH M. Critical comments on Hibler's presentation of his work on negative afterimages of hypnotically induced hallucinated colors. *J. exp. Psychol.*, 1941, **29**, 164–170.
3. ERICKSON, M. H., & ERICKSON, ELIZABETH M. Concerning the nature and character of posthypnotic behavior. *J. gen. Psychol.*, 1941, **24**, 95–133.
4. ESTABROOKS, G. H. *Hypnotism.* New York: Dutton, 1948.
5. KELLOGG, E. R. Duration of the effects of posthypnotic suggestion. *J. exp. Psychol.*, 1929, **12**, 502–514.
6. MOLL, A. *Hypnotism.* New York: Scribner & Welford, 1890.
7. PATTEN, E. F. The duration of posthypnotic suggestion. *J. abnorm. soc. Psychol.*, 1930, **25**, 319–334.
8. WEITZENHOFFER, A M. A note on the persistence of hypnotic suggestion. *J. abnorm. soc. Psychol.*, 1950, **45**, 160–162.

Received June 8, 1953.

SELF-PERCEPTION, PERCEPTUAL DEFENSE, AND ADJUSTMENT[1]

BERNARD CHODORKOFF

University of Wisconsin[2]

IN THE personality theories of Rogers (12, 13) and Snygg and Combs (2, 14), defensiveness is described as primarily a perceptual phenomenon which follows as a consequence of threat to the individual's self. Defense, in essence, is the prevention of accurate perceptions of what is threatening from reaching awareness. As a result, aspects of the environment and of the person himself may be denied to awareness or may be misperceived. It is in this way that the individual insures the stability of his self. Furthermore, the adequacy of the individual's personal adjustment is considered to be inversely related to the degree to which experiences are denied awareness.

The present research was designed to study the implication that the degree of defensiveness and its perceptual consequences will be the same, regardless of whether you select for examination the individual's perception of his environment or his perception of himself, and that both will be related to the adequacy of the individual's personal adjustment.

The hypotheses which were set up to be tested may be stated as follows:

1. The greater the agreement between the individual's self-description and an objective description of him, the less perceptual defense he will show.[3]

2. The greater the agreement between the individual's self-description and an objective description of him, the more adequate will be his personal adjustment.

3. The more adequate the personal adjustment of the individual, the less perceptual defense he will show.

METHOD

The methodology employed is a blend of clinical and experimental techniques. This approach made it possible to obtain relevant information about several currently important areas of research; these include perception and personality theory, perceptual defense, the role of perception in adjustment, and the relationship between self-perception and adjustment.

Instruments. Defensiveness in perception of the self was studied by means of a *Q* sort of self-descriptive statements. Disagreement between the individual's *Q*-sort self-description and the *Q*-sort description of him based on projective material represented defensiveness in this situation.

Defensiveness in terms of perception of the environment was studied by means of the tachistoscopic presentation of neutral words and threatening words for which the threat is personally relevant. In this situation defensiveness manifested itself as greater difficulty in recognizing the threatening words.

An adjustment rating scale and the Munroe Inspection Rorschach Check List (9) were used to evaluate the projective material. Two measures of adjustment were derived from these.

Subjects. The subjects (*S*s) were 30 male undergraduates who were taking introductory psychology at the University of Wisconsin. Their ages ranged from 18 to 29 years; mean age was 21.4 years with a standard deviation of 1.92 years. None of the *S*s had previously taken any of the tests used here, and none had ever been or was currently involved in psychotherapy or counseling.

Procedure. Three sessions were spent with each *S*. The first two sessions were scheduled as close together as possible, and the third was held exactly one week following the second session.

During the first session each *S* filled out a biographical inventory, and the Rorschach was administered. The biographical inventory consisted of 45 multiple-choice and incomplete-sentence-type items. The items were adapted from an inventory described by Kelly and Fiske (5). They provided background information about each *S*.

The second session included the administration of a word association test[4] and the Thematic Apperception Test. The word association test consisted of 100 words,

[1] This article is based on a thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of Wisconsin. The author expresses his gratitude to Dr. E. E. Baughman for his supervision and assistance, and to the other members of his research committee, Drs. D. A. Grant and J. C. Gilchrist, for their suggestions. The author also expresses his appreciation to the judges who participated in evaluating the projective material—Drs. S. Epstein and H. A. Page and Miss E. Whitehouse.

[2] Now at the Veterans Administration Hospital, Dearborn, Michigan.

[3] Perceptual defense here refers to the differential recognition of threatening and nonthreatening words.

[4] The following have been deposited with the American Documentation Institute: (*a*) the word association test—Document No. 4163; (*b*) the *Q* sort—Document No. 4162; (*c*) the adjustment rating scale—Document No. 4161. Order by document number from the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance for each one $1.25 for 35 mm. microfilm or $1.25 for 6 by 8 in. photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

50 emotional and 50 neutral. These were five-letter words selected on the basis of (*a*) equivalent frequency of usage as determined by the Thorndike-Lorge (16) word counts and (*b*) agreement between four judges as to the classification of each word. The *S*'s reaction time to each word was measured by means of a voice-key type of reaction time apparatus.

The Rorschach test and the TAT were administered and treated according to the procedures outlined by Klopfer and Kelley (6) and Murray (10).

The perceptual defense procedure was administered one week after the word association test. On the basis of the reaction times obtained on the word association test, two groups of ten words each were selected for each *S* to be presented as the stimuli for the perceptual defense procedure. One group consisted of the emotional words that had the highest reaction times; the other group consisted of the neutral words with the lowest reaction times. In no case was a word selected if *S* gave indication of not knowing its meaning. In this way it was possible to select for each *S* those words which represented *personally relevant threatening and neutral stimuli*.

Three practice words were presented tachistoscopically to *S*, and these were followed by the 20 selected words presented in random order. Each word was presented in the tachistoscope beginning at the .01-sec. exposure level, and the exposure time was increased consecutively by .01 second until *S* was able to give two consecutive, correct reports of the stimulus word. The exposure time at which he correctly reported the word for the second time was used as the recognition score for that word.

Immediately following the perceptual defense procedure *S* was given the *Q*-sort cards with instructions to sort them so that they described himself. This *Q* sort consisted of 125 short self-descriptive statements,[5] e.g., "I am impulsive"; "I usually ask for help if I can get it."

It was *E*'s purpose to have the *Q* sort contain items about which information could most readily be obtained from the Rorschach and TAT, and which would describe most people to some degree. An attempt was made to have the final sample follow the principle of randomization (15); that is, the items should interact in *S*'s judgment about them, but each item should not be contingent upon any other one. This final *Q* sort was developed with the cooperation of four clinically trained persons.

The following *Q*-sort distribution was employed:

Frequency:	3	5	7	10	13	16	17	16	13	10	7	5	3	
Scale value:	12	11	10	9	8	7	6	5	4	3	2	1	0	

$N = 125$

Three judges and *E* participated in the task of evaluating the projective and biographical material obtained from *S*s. The judges were selected on the basis of their clinical experience and ability to deal with projective data. Each judge worked on ten cases; these were assigned in consecutive order as the material became available. The *E*, as the fourth judge, teamed up with each of the other judges in evaluating the material, and, therefore, participated in evaluating all 30 *S*s.

[5] See footnote 4.

At the end of the second and before the third session, one of the judges was given the following material: the biographical inventory, the Rorschach protocols with the scores summarized, the word association test data, and the TAT protocols.

The judge and *E* worked up their independent interpretations of the data, and then met to discuss them. After the two reached some agreement about their evaluation of the material, they sorted the *Q*-sort items to describe *S* as they saw him.[6]

The judge and *E* independently rated *S* on an adjustment rating scale[7] which consisted of 11 subscales. A final rating for *S* on each of the subscales was then agreed upon. The content of each scale was derived primarily from Maslow and Mittleman (8, ch. IV). A descriptive paragraph was provided for the extreme ratings of each scale so that these might serve as anchoring points for the judges.

After the judges completed their work, *E* applied the Munroe Inspection Rorschach Check List to *S*s' Rorschach protocols.

Measures and statistics. Three basic scores were derived from the data. The first was the accuracy of self-description (ASD) score, which was obtained in the following way:

The judges' *Q*-sort description of *S* was considered the objective description of him. The *S*'s *Q*-sort self-description was correlated with the judges' *Q*-sort description. The product-moment r obtained was transformed to Fisher's z, and this score was used as the ASD score. The higher the ASD score, the greater the agreement between the objective description of *S* and *S*'s self-description.

The second basic score was the perceptual defense (PD) score. Here two scores were derived. One was the difference between the mean recognition threshold for neutral words and the mean recognition threshold for threatening words (Mt − Mn); this was called the PDm score. The other was the difference between the standard deviation of the distribution of recognition thresholds for the neutral words and the standard deviation of the distribution of recognition thresholds for threatening words ($\sigma t - \sigma n$); this was called the PDs score.[8]

The third score was the adjustment (A) score. Here again, two measures were derived, one based on the Munroe check list and the other derived from the judges' ratings of adjustment. The check-list adjustment (Ac) score was the number of checks *S* received.

[6] This procedure was suggested by Dr. Julius Seeman, Counseling Center, University of Chicago.

[7] See footnote 4.

[8] In the present study it was hoped that the possibility of the occurrence of response suppression in the perceptual defense situation, and intentionally produced inaccurate self-description in the *Q*-sorting procedure, would be minimized by using only male *S*s and by conducting the experimental procedures in an atmosphere which would not inhibit *S*s. It is doubtful that these did occur to any significant degree especially since the sequence of the procedures would also tend to lessen the possibility of inhibition. However, if these reactions did occur, then these should also be considered as defensive reactions.

The rating scale adjustment (Ar) score was derived by converting each *S*'s rating on each scale into a standard score based on the distribution of the 30 ratings for that scale. For each *S*, the Ar score was the sum of his standard scores on the 11 subscales. Poor adjustment was indicated by a high check-list score, or by a low rating on the adjustment rating scale.

RESULTS

Table 1 presents the correlations that were obtained between the basic scores. It will be noticed in this table that in addition to product-moment correlations, biserial and multiple correlations were also employed. Product-moment correlations were not used with the Ar scores because these scores were based on values obtained from a rating scale on which the equality of units was questionable. The group of 30 *S*s was divided into good and poor adjustment groups by using the median as the dividing point. Biserial r was then appropriate (3). Multiple correlations were used because of the possibility that a combination of the two measures of adjustment would give a better estimate of *S*'s adjustment level than would either measure alone.

The findings are listed below as they relate to their respective hypotheses:

Hypothesis 1. Both the PDm and PDs scores correlated negatively and significantly with the ASD scores. These findings support the first hypothesis. The more accurately the individual described himself, the less perceptual defense he showed.

Hypothesis 2. The Ar scores correlated positively and significantly with the ASD scores. Although the Ac scores did not correlate significantly with the ASD scores, the correlation nevertheless approached significance in the predicted direction. A combination of the Ar and the Ac scores correlated positively and significantly with the ASD scores. In general, these results support the second hypothesis. The better the adjustment of the individual, the more accurate was his self-description.

Hypothesis 3. The Ar scores correlated negatively and significantly with the PDm scores. The correlation between the Ac and PDm scores, although not significant, was in the predicted direction. A combination of the Ar and the Ac scores correlated negatively and significantly with the PDm scores.

The Ar scores correlated negatively and significantly with the PDs scores. The Ac scores correlated positively and significantly with the PDs scores. A combination of the Ar and the Ac scores correlated negatively and significantly with the PDs scores. These results confirm the third hypothesis. The more adequate the individual's personal adjustment, the less perceptual defense he showed.

In addition to the correlations listed in Table 1, the two measures of adjustment were correlated. A biserial r of $-.56$ was obtained and is significant beyond the .05 level.

It should be noted that the adjustment scores that were used were not independently obtained criterion measures. The judges' *Q*-sort descriptions of *S*s were based on the same projective material from which both measures of adjustment were derived. It would have been advantageous to have obtained an independent evaluation of adjustment, but unfortunately this was not feasible. Consequently, the possibility that the correlations of the Ar and Ac scores with the ASD scores are somewhat inflated must be considered. The correlations involving the PD scores are not open to the same criticism, however.

TABLE 1

SUMMARY OF CORRELATIONS

Hypothesis	Basic Scores Correlated	Type of Correlation	Value
1	ASD with PDm	r	$-.53^{**}$
	ASD with PDs	r	$-.44^{*}$
2	Ac with ASD	r	$-.34$
	Ar with ASD	r_b	$.73^{**}$
	Ar and Ac with ASD	R	$.73^{**}$
3	Ac with PDm	r	.29
	Ac with PDs	r	$.37^{*}$
	Ar with PDm	r_b	$-.62^{**}$
	Ar with PDs	r_b	$-.68^{**}$
	Ar and Ac with PDm	R	$-.62^{**}$
	Ar and Ac with PDs	R	$-.68^{**}$

Note.—Product-moment correlation has been indicated by r. Biserial correlation has been indicated by r_b and multiple correlation by R.

* Significant at the .05 level of confidence.

** Significant at the .01 level of confidence.

DISCUSSION

Adjustment and the handling of perceived threat. An interesting finding became evident when the individual scores were perused. Of the 30 *S*s, 11 recognized the threatening words more quickly than the neutral words. These *S*s appear to be the sensitizers spoken about in other perceptual defense studies (1). Of these 11 *S*s, 10 belong to the group of 15 *S*s

who have the highest ASD scores. Also, of these 11 sensitizers, 8 are included in the group of 15 *S*s who have the highest Ar scores. When these results are considered with the major findings of this study, the following formulation is suggested:

The better adjusted individual attempts to obtain mastery over threatening situations by getting to know, as quickly as possible, what it is that is threatening. He may be alerted to the possibility of threat and may be influenced by a set which leads him to try to differentiate and symbolize aspects of himself and his environment which are potentially threatening. The less adequately adjusted individual, in contrast, may be thought of as influenced by a set to keep threat inadequately differentiated and inadequately symbolized.

Furthermore, it can be argued that adequate behavior depends upon adequate perception. If an individual cannot differentiate and symbolize aspects of himself and his environment, his behavior in turn will not be adequate. He will also be unable to resolve threatening situations when they confront him, for if he does not recognize threat, how can he be expected to deal with it effectively?

It may be that for the well-adjusted person threat is a relatively momentary phenomenon. He is only threatened initially; then he extends himself in order to handle the threat. As a result, the threat is minimized. For the maladjusted person, threat persists because he cannot deal with it directly. Instead he tries to remove himself from it by denial or misperception.

Perceptual defense an individual, not a general, reaction. The concept of *perceptual defense*, an unconscious mechanism of resistance to recognition of threatening material, has frequently been used to explain the higher recognition thresholds for taboo words than for neutral words (1, 7). Whether or not such a mechanism is necessary has been questioned by some workers in the field (4, 11).

In the present study, no difference was found between the mean recognition thresholds for threatening and nonthreatening words. The t test was used to evaluate the difference between the means (Mt = 15.11, σt = 7.62; Mn = 14.35, σn = 8.91), and a value of t = .75 was obtained. This is obviously not significant and indicates that perceptual defense did not occur for the group as a whole. The matter cannot be dismissed at that, however, for meaningful differences have been found in terms of individual reactions. The importance of these differences has been indicated by the major findings of this study. These findings suggest therefore that perceptual defense is not a general phenomenon but may be a process which some individuals are more apt to manifest than others.

Importance of using individually selected, personally relevant stimuli. Consideration must be given to the possibility that there are different kinds of threat. Taboo words like *whore* and *bitch* may present threat in the sense of social prohibition which is uniformly imposed on most people; these words therefore, can generally be expected to produce elevated recognition thresholds. Other words like *mammy* and *haven*, if threatening to the individual, may represent more basic psychologic threat, and may be more intimately related to the person's personality trends. The handling of such threatening material is much more of an individual matter.

SUMMARY

The present research was designed as an attempt to verify the following hypotheses:

1. The greater the agreement between the individual's self-description and an objective description of him, the less perceptual defense he will show.

2. The greater the agreement between the individual's self-description and an objective description of him, the more adequate will be his personal adjustment.

3. The more adequate the personal adjustment of the individual, the less perceptual defense he will show.

The results in general supported all hypotheses. The following can be concluded:

1. In a group of *S*s who show varying degrees of adjustment and defensiveness, one finds that the more inaccurate and faulty the individual's perception of his environment, the more inaccurate and faulty is his perception of himself; and the more inaccurate and faulty the individual's perceptions of himself and his environment, the more inadequate is his personal adjustment.

2. Perceptual defense is meaningful when it is considered as an individual phenomenon rather than a general or universal one. Important differences exist between persons and

apparently can be demonstrated if personally relevant, individually selected stimuli are used.

REFERENCES

1. BRUNER, J. S., & POSTMAN, L. Emotional selectivity in perception and reaction. *J. Pers.*, 1947, **56**, 69–77.
2. COMBS, A. W. A phenomenological approach to adjustment theory. *J. abnorm. soc. Psychol.*, 1949, **44**, 29–35.
3. CRONBACH, L. J. Statistical methods applied to Rorschach scores: a review. *Psychol. Bull.*, 1949, **46**, 393–429.
4. HOWES, D. H. & SOLOMON, R. L. A note on McGinnies' "Emotionality and perceptual defense." *Psychol. Rev.*, 1950, **57**, 229–234.
5. KELLY, E. L., & FISKE, D. W. *The prediction of performance in clinical psychology*. Ann Arbor: Univer. of Michigan Press, 1952.
6. KLOPFER, B., & KELLEY, D. M. *The Rorschach technique*. Yonkers, N. Y.: World Book Co., 1946.
7. MCGINNIES, E. Emotionality and perceptual defense. *Psychol. Rev.*, 1949, **56**, 244–251.
8. MASLOW, A. H., & MITTLEMAN, B. *Principles of abnormal psychology*. New York: Harper, 1941.
9. MUNROE, RUTH L. Prediction of adjustment and academic performance of college students by a modification of the Rorschach method. *Appl. Psychol. Monogr.*, 1945, No. 7.
10. MURRAY, H. A. *Thematic Apperception Test*. Cambridge: Harvard Univer. Press, 1943.
11. POSTMAN, L., BRONSON, W. C., & GROPPER, G. L. Is there a mechanism of perceptual defense? *J. abnorm. soc. Psychol.*, 1953, **48**, 215–223.
12. ROGERS, C. R. The organization of personality *Amer. Psychologist*, 1947, **2**, 358–368.
13. ROGERS, C. R. *Client-centered therapy*. New York: Houghton Mifflin, 1951.
14. SNYGG, D., & COMBS, A. W. *Individual behavior: a new frame of reference for psychology*. New York: Harper, 1949.
15. STEPHENSON, W. Some observations on *Q* technique. *Psychol. Bull.*, 1952, **49**, 483–498.
16. THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Bureau of Publications, Teachers Coll., Columbia Univer., 1944.

Received June 25, 1953.

AN EXPERIMENTAL INVESTIGATION OF THE EFFECT OF UNSTABLE INTERPERSONAL RELATIONS IN A GROUP[1]

LEON FESTINGER

Laboratory for Research in Social Relations, University of Minnesota

AND HERMAN A. HUTTE

Nederlands Instituut voor Praeventieve Geneeskunde, Leiden, Netherlands

IN ORDER to work together, live together, or even to maintain the barest degree of social contact with one another, it is necessary for persons to share ideas, to communicate attitudes and feelings and information. While much communication is intended as such, some of the communication which occurs among persons may be termed "indirect" or "inadvertent" communication. This occurs when information is conveyed to others, not by actually talking about it, but by behaving in such a way that someone else can draw inferences.

There are some kinds of things about which communication is almost never direct but is typically inadvertent. Among adult persons of the same sex, for example, it is quite rare to have direct communication about how much they like each other. This type of direct communication about social preferences does go on among children, but among adults it would generally be embarrassing. People do, however, behave differently toward those they like and toward those they dislike and in this manner achieve communication about these feelings. Certainly, among people who interact with one another, there is knowledge of how others feel about them even though there may never have been any direct communication on the subject.

How much knowledge persons in a group have concerning how others feel about them is determined by two factors, assuming a complete absence of any direct communication on the subject. One factor is the amount of difference between behavior toward those who are liked and toward those who are disliked. The other factor is how sensitive others are to these differences in behavior. In a group that behaves very formally and strictly according to some rules of etiquette and politeness, for example, we would expect the differences in behavior toward those liked and disliked to be minimal. We would not, consequently, expect to find very accurate knowledge in such a group concerning likes and dislikes. Also, of course, no matter how large the differences in behavior directed toward different persons might be, we would expect little accurate knowledge if, for one reason or another, people paid no attention to, or were not sensitive about, these cues.

This latter factor of sensitivity is maximally important in inadvertent communication and minimally important in direct communication. In direct communication the person must only be listening in order to have the communication impinge upon him. In inadvertent communication, however, this factor of sensitivity may make a great difference.

The present experiment attempts to study the effect of the stability of the interpersonal relations within the group on the accuracy of perception of social preference. It may be expected that persons whose relations with others in a group are stable may be different from those whose relations are unstable with respect to both sensitivity and the cues they make available concerning their feelings for others.

It is possible that those whose relations in a group are stable, because of their greater security, would be less cautious in their behavior and, consequently, provide more cues from which others could infer their likes

[1] The collaboration on this study was made possible by a grant from the Rockefeller Foundation. Initial planning of the study was carried out jointly by the two authors in Ann Arbor, Michigan. After preliminary experimentation was done, the final design was agreed upon during conferences in Leiden, Holland. The actual experimental work in Holland was done at the Nederlands Instituut voor Praeventieve Geneeskunde with funds provided by the Nederlandse Organisatie voor Zuiver Wetenschappelijk Onderzoek. In the United States it was done at and supported by the Laboratory for Research in Social Relations of the University of Minnesota. The authors wish to express their appreciation and indebtedness to Aart Hazewinkel and Els v. d. Schoot, who helped conduct the experiment in Holland, and to Jay Jackson and Robert Hall, who helped in the United States.

and dislikes. One might also expect that those whose relations in the group are unstable would be more sensitive to the cues available because they would be more concerned with finding out how others feel toward them. The experiment was consequently designed so that half the people in each group were put in a situation where they felt secure and stable about their interpersonal relations in the group while the other half were made to feel unstable about these relations. After they had had a period of time in which to interact with one another, measurements were made from which one can calculate how accurate their knowledge was of how others felt about them. From observation of the interaction among them one can calculate whether there was any difference between *S*s in the two conditions in behavior toward others.

The experiment was conducted simultaneously in two different countries, Holland and the United States. The same number of groups was run and the identical procedures were used (within the limits of translation of language between English and Dutch) in both countries. The comparison of the results in the two countries can provide us with some evidence concerning how independent the findings are of cultural differences, at least between these two countries.

METHOD

Subjects. In each of the two countries data were collected on 40 experimental groups, each group consisting of six subjects (*S*s) of the same sex. Twenty male and 20 female groups were run in Minneapolis; 28 male and 12 female groups were run in Holland. All *S*s were undergraduate students at the University of Minnesota, Leiden, or Amsterdam. The *S*s scheduled for any particular group were always strangers to one another. For the purposes of this experiment stranger was defined as not knowing the other person's name. Occasionally there were persons in the same group who knew each other by sight but did not know each other's name. In Holland there were occasional deviations from this rule. Persons who knew each other's name were sometimes in the same group if the experimenter (*E*) felt sure that their acquaintance was only superficial.

Procedure. The *S*s, upon their arrival, were seated in a U-shaped arrangement. In the United States the observers (*O*s) were located behind a one-way observation mirror which faced the open end of the U. In Holland one *O* was sometimes behind a one-way observation mirror and sometimes in the room with *S*s. The other *O*s were always in the room with *S*s. In this way it was possible for *O*s to see all the faces of the persons in the group. A letter from A to F was placed in front of each *S* for identification purposes.

The *S*s were asked to introduce themselves to each other by telling their names, what part of the country they came from, and what they were studying. This was the only information any of the *S*s had about the others at the start of the experiment.

The *E* then verbally gave the following instructions:

> This afternoon you will have a discussion among the six of you in which you will make up a plot for a radio play. Before we start doing that it is, however, necessary for the purposes of this experiment for us to know how you feel about each other. After the experiment I will explain to you why we do this. We would like you to write down on a slip of paper which will be passed out to you in a minute how well you like each of the other persons in this group. On this slip we would like you to write down the letters of each of the other persons in your group alongside that statement that best expresses your feelings toward each other.

The following slip was then passed out:

My Letter is ____

7. likes me exceptionally well _____
6. likes me quite well _____
5. likes me moderately _____
4. likes me a little _____
3. rather indifferent _____
2. dislikes me a little _____
1. dislikes me definitely _____

As soon as the slips had been passed out, *E* continued:

> Some of you may be a bit doubtful about these "first impressions" of people. I can, however, assure you that we have found in the past, in many studies we have done, that first impressions of people are extremely important and surprisingly valid. For example, it frequently happens that spontaneous lasting friendships are formed which correspond exactly to such first impressions. Also, spontaneous dislikes sometimes occur between people who met for the first time which prove to be quite stable. We would like you, then, to take your time in filling out these slips so that you can do it seriously.

The last instruction was given in an attempt to make *S*s take these initial ratings seriously. Since the adequacy of the experimental manipulation depended to a great extent on how important *S*s felt these initial ratings to be, this last instruction was put quite forcefully. After *S*s had finished filling out the slips, they were also asked to encircle the letters of the two persons in the group whom they liked best. The slips were then collected and *E* continued:

> We would now like you to estimate how each of the others rated you a moment ago. For that purpose a slip of paper will now be passed out to you. Will you put the letter of each other person in the group alongside that statement which describes the way you think that person feels toward you.

After this second slip, similar in form to the first slip, had been filled out and collected, *E* went on:

> We will now give you some information about how the two that you like best in the

> group feel about each other. It is possible to give you this information because while you were filling out the second slips, Mr.———— has been tabulating the first slips. We will now pass out to each of you a slip which will tell you how those two which you said you liked best feel about each other.

The slips containing this information were then distributed. The actual information given to *S*s at this point was fictitious and constitutes the experimental manipulation. We will describe this in detail in the next section. The *E* then said:

> We are now ready for you to start your discussion about the radio play. We will pass out to you a plot which is unfinished and we would like you to discuss among yourselves how you think it should end. You are completely free to decide upon any solution of the problems in the plot. Keep in mind a short scene lasting about ten minutes. I will now pass out copies of the unfinished plot to each of you. Please read it carefully and let me know when you are ready to start the discussion.

The following plot was then passed out:

> Henry, the son of a physician, has a friend, Jim, who is under the care of Henry's father. Henry knows that Jim is incurably ill. Both are in love with a girl, Ellen. Jim doesn't know what kind of disease he has; neither does Ellen know he is incurably ill. One night Henry calls on Ellen just after he had decided to give up his studies and accept a job in California. He intends to ask her that night to marry him and go with him to California. Henry knows that for many years Ellen has wanted to go to live there. Before he gets a chance, however, to tell her, Ellen announces her engagement to Jim. What is Henry going to say and do and what will happen?

After all six *S*s had finished reading the plot, *E* started the discussion, telling *S*s that he himself would sit at the side of the room and would not participate. The discussion went on for 20 minutes during which time *O*s recorded certain aspects of the communication process. These observations will be described in detail later. When 20 minutes had gone by from the time the discussion started, *E* said:

> We now have to interrupt your discussion for a moment because we want you to repeat what you did at the beginning, that is, to indicate again on the slips of paper to be passed out to you how you feel about the others in the group.

Slips identical in form to the first slips were then passed out. After they were filled out by *S*s and collected, *E* continued:

> And we also would like you to again estimate how the others in the group rated you. You see, one of the main purposes of the experiment is to see how likes and dislikes are formed and how well people know how others feel about them. So please give this your full attention and think it over well before you write down your estimates.

After these slips had been distributed, filled out, and collected, *S*s were given a questionnaire to answer. We will describe the exact questions asked as we deal with them in discussing the results. The *S*s were then told that the experiment was over and were given a full explanation of what had been done and why.

Experimental manipulation of stability of interpersonal relations. When *S*s were given information as to how the two persons they liked best[3] felt about each other, this information was systematically fabricated to produce conditions of stability or instability of interpersonal relations. In each group, three *S*s were told that their two top choices liked each other (one rating of 6 and one of 7), while the other three were told that their top choices disliked each other (one rating of 1 and one of 2). The former information was intended to make *S*s feel that their relations in the group were comfortable in that those persons whom they liked best were also compatible with one another. This condition, which we will call the compatible condition, should have the effect of stabilizing their preferences for others and should make them feel relatively secure about their personal relations. The condition in which *S*s were told that their top choices disliked each other was intended to have the opposite effect. Here it was felt that the knowledge of the incompatibility between the two people they liked best would make them feel uncomfortable about their relations with others in the group. This condition, which we will call the incompatible condition, should make *S*s relatively uncertain about their interpersonal relations in the group and should make their preferences for others unstable. Such relations among three persons have also been discussed by Heider (1) in a manner similar to the present discussion. We will have more to say later concerning the relevance of some of our findings to Heider's theory.

The *S*s were put at random in either the compatible or the incompatible condition. Prearranged sequences were followed which systematically varied condition by position around the table. The *S*s were in one or another condition depending upon the particular position sequence predesignated for that group.

Observation of the group discussion. During the 20-minute discussion, observations were made of who talked, for how long, and to whom. In both countries the recording of the observations was done on a moving tape so that the duration of any communication could be computed afterwards. The *O*s had to indicate which person spoke and when he stopped speaking. Direction of communication was indicated by either writing down the letter of the person addressed or indicating, where appropriate, that the remark had been addressed to the group as a whole.

For the data collected in Leiden, each group was observed by three persons. The records of one of the *O*s was used for the analysis. His record was modified if both other *O*s disagreed with him. The reliability of observation of who spoke to whom was .84 computed by product-moment correlation. For the data collected in Minneapolis, two *O*s were present for only 10 of the 40 groups. On 66 per cent of all items recorded the two *O*s agreed both as to who was talking and to whom the

[3] For the sake of ease of expression, we will refer to the two persons who are most liked as the two top choices.

communication was directed. Most of the disagreements involved one *O*'s saying that the remark was directed to some specific person while the other recorded it as having been directed toward the group. The data used in analysis are those observations made by that *O* who observed all of the groups.

Questionnaire data. Data collected on slips which *S*s filled out during the experiment and also data collected on the postexperimental questionnaire can both be considered questionnaire data. There follows a summary of the data thus collected:

1. Ratings of how well *S*s like the other members of the group both before and after the discussion.
2. Estimates of how well each of the others like them both before and after the discussion.
3. Estimates of how much *S*s' top choices like each other after the discussion.
4. Perceptions of how much *S*s talked to others and others talked to them.

RESULTS

Adequacy of the experimental manipulation. It will be recalled that the experimental manipulation consisted of an attempt to have some persons think that the two members of their group whom they liked best *liked each other* (compatible condition), while others were given the impression that their two top choices *disliked each other* (incompatible condition). In order to demonstrate how adequately this was accomplished, it is not sufficient to show that *S*s believed what they were told. The course of the experiment, including the discussion, might present other evidence to them. The adequacy of the manipulation can only be judged in terms of whether, at the conclusion of the experiment, they still retained the impression which they were given.

In order to gather these data, *S*s were asked in the postexperiment questionnaire to state once more which two members of the group they had liked best at first. They were then asked, "What is your best judgment of how those two feel about each other now?" They were given an opportunity to check one of four possible answers on a scale from (1) they dislike each other to (4) they like each other very much. Table 1 presents the results of the answers to this question.

TABLE 1

JUDGMENTS OF HOW WELL THE TWO INITIALLY BEST LIKED PERSONS LIKE EACH OTHER AFTER THE EXPERIMENT

Condition	Minneapolis		Leiden	
	N	Mean Rating*	*N*	Mean Rating*
Compatible	120	3.4	120	3.3
Incompatible	118†	2.1	120	1.9

* Scale from 1 (they dislike each other) to 4 (they like each other very much).

† Two *S*s did not answer the question.

It is clear that in both countries the manipulation was successful. The *S*s in the compatible condition judged that their top choices like each other much more than did *S*s in the incompatible condition. In every one of the 80 groups the average for *S*s in the compatible condition was higher than for *S*s in the incompatible condition. We may then conclude that the difference is highly significant. It is also interesting to note that the mean values for each of the two conditions are very much alike for both countries.

Let us now turn our consideration to the indirect effects which it was hoped the manipulation would have. Our hypothesis was that *S*s in the incompatible condition would be less stable in their liking of others than *S*s in the compatible condition. Because they had been given the impression that their top choices liked each other, *S*s in the compatible condition would find themselves in a stable interpersonal situation, and we would not expect many of them to change their preferences during the experiment. On the other hand, *S*s in the incompatible condition who had the impression that their top choices disliked each other, would feel uncomfortable about the interpersonal situation, and we would expect many of them to change their feelings about at least one of their two best-liked persons during the experiment. Of course, there are many other factors aside from this one that could induce people to change their preferences, and so, if the above hypothesis is correct, we can only hope to demonstrate a difference between the conditions. Table 2 presents the data on the number of persons in each condition who changed at least one of their top two choices.

It can be seen that those in the incompatible condition do indeed change at least one of their two top choices much more frequently than do those in the compatible condition. Again, the data from both countries are very similar although there is a slightly larger difference in the Leiden data than in the Minneapolis data. When we evaluate the signifi-

TABLE 2
NUMBER OF PERSONS CHANGING TWO TOP CHOICES DURING EXPERIMENT

Condition	Number Changing at Least One of Their Two Top Choices		Number not Changing	
	Mpls.	Leiden	Mpls.	Leiden
Compatible	66	53	54	67
Incompatible	96	94	22	26

cance of the differences on the basis of consistency from group to group, we find the differences to be highly significant. The number of groups in which more incompatibles than compatibles changed at least one of their two top choices is 25 and 26 for Minneapolis and Leiden respectively. The number of groups in which the direction of difference was reversed was 7 and 4 for Minneapolis and Leiden respectively. The same number changed in the other 18 groups. Tested by chi square (ignoring the equal groups) the difference is significant beyond the .01 level of confidence in each country.

The data do corroborate the hypothesis, and we may conclude that in both countries, one of the effects of the manipulation was to produce a tendency for the incompatibles to change at least one of their two top choices during the experiment. This result is also corroboration of Heider's (2, 3) theory concerning such relationships.

If it is true that many *S*s in the incompatible condition are changing their two top choices because of the perception that those two dislike each other, we would expect them to perceive greater liking between those they finally choose as the ones they like best. Among the *S*s in the compatible condition, however, changes which occurred would have been for any of a variety of other reasons. We would then expect no such change in perception of how the two whom they liked best after the experiment like each other.

In the questionnaire at the conclusion of the experiment *S*s were asked to state which two in the group they *now* liked best and also to state their best judgment of how these two now felt about each other. These data are presented in Table 3.

For those in the compatible condition, the changes in average perceived liking between their two top choices are negligible. For *S*s in the incompatible condition the shifts in their two top choices are significantly toward persons who are perceived as liking each other better. Looking at consistency from group to group we find that in 32 of the Minneapolis groups and in 33 of the Leiden groups the change in perceived compatibility from the initial top choices to the final top choices was greater in the incompatible than in the compatible condition. In only four groups (all in Minneapolis) was the direction reversed. We thus have confirmation that, in the incompatible condition the experimentally produced belief that top choices did not like each other was a major factor in producing changes in preferences.

The relation between social preference and frequency of communication. During the course of the 20-minute discussion, observations were made of who spoke, for how long, and to whom the communications were directed. We may look at these data to determine whether the direction of communication was affected by the preferences which persons had for the others in the group. To the extent that there are differences in how often members spoke to those they liked, to those they disliked, or to those they were indifferent to, we may say that cues were present during the discussion from which these preferences could be judged. The data on the average length of each communication do not show any clear trends. Our presentation will consequently deal with the frequency of communication, that is, how often communications were addressed to certain others in the group.

In examining the relation between preferences for others and number of communications addressed to them, it is possible to look

TABLE 3
PERCEIVED PREFERENCES BETWEEN TWO TOP CHOICES BY THOSE WHO CHANGE AT LEAST ONE OF THE TOP CHOICES

Condition	Judgment of How Well Two Top Choices Like Each Other			
	Minneapolis		Leiden	
	Initial Two	Final Two	Initial Two	Final Two
Compatible	3.1	3.2	3.1	2.9
Incompatible	2.1	2.9	1.8	2.8

at the data in relation to initial preferences (ratings made at the beginning of the discussion) or in relation to the final preferences (ratings made after the discussion). In general, the same trends are reflected in both ways of examining the data. Since our primary objective is to determine whether or not cues were present from which members of the group could judge how others felt about them, and since the accuracy of this judgment is in terms of the preferences stated after the discussion, it makes more sense to look at the relation between frequency of communication and final preferences. We shall, therefore, confine our discussion here to this relationship. Table 4 presents the relevant data.

There is clearly a relationship between the rating given to a person (how much one likes him) and the number of communications addressed to that person. But this relationship is not monotonic. For both the compatible and incompatible conditions, both in Minneapolis and in Leiden, the same type of relationship is obtained. The fewest number of communications are addressed to those rated 3 (indifferent). The frequency of communication to those rated 1 or 2 (dislike) is considerably greater, and it also increases on the other side of the continuum. More and more communications are addressed to people the better they are liked. The differences between the frequency of communication to those rated 3 and to those rated 6 or 7 are significant at at least the .05 level for each condition in each country. The differences between the number of communications addressed to those rated 3 and to those rated 1 or 2 are also significant at beyond the .05 level for each condition in each country. We may conclude that the amount of communication was an available cue which could be used by the members of the group to judge how others felt about them and that these cues were the same, roughly, in both countries. It is likely that the communications addressed to those rated 1 or 2 were qualitatively different from those addressed to people rated 6 or 7 but, unfortunately, no data were collected which are relevant to this point.

TABLE 4

MEAN NUMBER OF COMMUNICATIONS ADDRESSED TO OTHERS ACCORDING TO FINAL RATINGS OF PREFERENCE

Condition	Mean Number of Communications Addressed to Those Rated					
	1 or 2	3	4	5	6	7
Compatible (M)	4.1	1.5	1.6	1.9	3.4	4.6
Compatible (L)	7.1	2.5	3.4	3.7	4.0	4.8
Incompatible (M)	3.0	1.4	2.2	3.2	3.7	5.7
Incompatible (L)	4.1	2.2	3.3	4.4	5.1	7.2

We may now raise the question as to whether or not the communication pattern, and hence the available cues, were different between the compatible and incompatible conditions. We may note in Table 4 that for both countries the frequency of communication addressed to those rated 1 or 2 is higher for the compatible condition than for the incompatible condition. The figures are about equal between the two conditions for those rated 3 (indifferent). In the incompatible condition *S*s communicated more to those rated 5, 6, or 7 than in the compatible condition. In other words, in the incompatible condition *S*s communicated relatively more to those they like and relatively less to those they dislike. While the consistency of the differences between the two countries lends weight to them, the differences within each country between the compatibles and incompatibles are not highly significant. For the Minneapolis data the difference between the conditions is significant ($p < .01$) for communications addressed to those rated 4 or 5. For the Leiden data the difference between conditions in frequency of communication to those rated 1 or 2 is significant ($p < .05$). None of the other differences approaches an adequate level of significance. Mainly because of the startling consistency between the two countries, we may conclude that there is some difference in the nature of the cues made available by those in different conditions.

We can make only tentative suggestions as to why this difference exists between the conditions. It is possible that *S*s in the compatible condition, feeling more secure in their personal relationships, are more able to directly express their likes and dislikes toward others. Thus they are able to react to those they dislike. The *S*s in the incompatible condition, on the other hand, may feel less able to express their dislikes directly and so, whenever possible, prefer to withdraw from

contact with those they dislike and maintain more contact with those they like. If this interpretation of the difference between conditions is correct, we would expect to find that *S*s in the compatible condition show their dislike when talking to those they rate 1 or 2. Unfortunately we do not have these data, and so the interpretation must remain highly tentative.

In comparing the data from Minneapolis with the data from Leiden one is struck by the consistently higher numbers of individual communications in the Leiden data. This is a function of the proportion of the communications that *O*s designated as directed toward the group rather than toward specific individuals. In Minneapolis, 58 per cent of the communications were recorded as directed toward specific individuals while in Leiden the comparable figure was 75 per cent. This was probably a difference in observer practice rather than a real difference in behavior of the groups. If we correct the Minneapolis or the Leiden data for the difference in proportion of communications judged to be addressed to specific individuals, the difference in absolute frequency disappears.

Perception of the pattern of communication. In the preceding section we have demonstrated that cues were available from the frequency of communication on the basis of which it would be possible for an individual *S* to judge how others in the group felt toward him. Before these cues could be used, of course, it would be necessary to perceive them. We will now examine data relevant to the question of whether or not these cues were perceived.

On the questionnaire after the experiment *S*s were asked to indicate on a scale from 0 (not at all) to 4 (very much) how much they had talked to each of the others in the group and also how much each of the others had talked to them. We may then examine these data to determine if *S*s were aware of the pattern of communication that existed in the group.

Table 5 presents the data on the relationship between *O*s' records of the discussion and the ratings on the questionnaire.

The data in Table 5 show clearly that *S*s were able to report the pattern of communication fairly accurately, both as to whom they talked to and who talked to them. Again the data are similar between the two conditions and between the two countries. It is important to note that, considering the data in a somewhat different light, Table 5 not only shows that *S*s perceived the pattern of communication, but also it corroborates the validity of the observation of direction of communication.

TABLE 5

ACCURACY OF PERCEPTION OF THE COMMUNICATION PATTERN

	Rating on Questionnaire of "How Much I Talked to Them"				
	0	1	2	3	4
Observed average frequency of communication given by subject in:					
Compatible Condition (M)	.5	1.4	2.3	5.0	7.4
Incompatible Condition (M)	.6	1.8	3.0	5.4	8.3
Compatible Condition (L)	.4	1.7	3.8	7.0	11.1
Incompatible Condition (L)	.6	2.0	4.1	7.6	13.0
	Rating on Questionnaire of "How Much They Talked to Me"				
	0	1	2	3	4
Average observed frequency of communication received by:					
Compatible Condition (M)	.5	1.4	2.3	4.5	7.0
Incompatible Condition (M)	.5	1.2	3.0	5.6	7.7
Compatible Condition (L)	.5	1.7	4.5	7.1	12.1
Incompatible Condition (L)	1.2	2.4	4.2	7.5	12.3

We may now ask whether or not the amount of communication directed toward specific persons, which potentially could give cues for judging how the other person felt, seems to have been used in this manner by *S*s. Table 6 presents the relation between the average number of communications received and *S*s' estimates of how well the communicator liked them.

Inspection of Table 6 shows that *S*s in both conditions and in both countries seem to have used the available cues. In all instances, the same nonmonotonic relationship that we observed in Table 4 holds. The *S*s judged those who talked to them least as being indifferent toward them. Those whom *S*s judged to either like or dislike them direct communications to them more frequently than those judged to be indifferent. In two out of the four comparisons

TABLE 6

RELATION BETWEEN AVERAGE FREQUENCY OF COMMUNICATION RECEIVED AND FINAL ESTIMATES OF HOW WELL OTHERS LIKE THEM

Condition	Final Estimates of Preference				
	1 or 2	3	4	5	6 or 7
Compatible (M)	3.9	2.1	2.1	2.7	4.0
Incompatible (M)	2.5	1.9	3.0	4.1	4.1
Compatible (L)	5.4	2.9	3.2	4.1	5.6
Incompatible (L)	5.5	2.3	3.5	5.2	6.0

(two conditions in each of two countries) the difference in amount of communication received from those judged to be indifferent (rating of 3) and those judged to dislike them (ratings of 1 or 2) is significant ($p < .05$). The difference between communication received from those judged to rate them 3 and those judged to rate them 6 or 7 is significant at the .01 level in every instance. We may then conclude that *S*s in both conditions, in both countries, seem to have used the cues, and we may then expect to find that they know, significantly better than chance, how others in the group feel toward them after the discussion.

Accuracy of knowledge of preference. We may now turn our attention to the question of whether or not the accuracy of estimation of how others felt about them was better than chance.

For each estimate that *S*s made concerning how others felt about them, we may calculate the average actual rating of preference received. If there are no differences among these average ratings for different estimates, then we would conclude that *S*s had no knowledge of how others really felt about them. Table 7 presents the data on accuracy of estimation of preference at the beginning of the experiment, that is, before *S*s knew each other or had had much opportunity to interact.

TABLE 7

ACCURACY OF ESTIMATION OF PREFERENCE BEFORE DISCUSSION

Condition	Estimate of Others' Preference for Themselves				
	1 or 2	3	4	5	6 or 7
Compatible (M)	5.1	4.7	4.7	4.6	4.9
Incompatible (M)	4.7	4.5	4.7	4.8	5.0
Compatible (L)	4.3	4.3	4.5	4.8	4.9
Incompatible (L)	4.2	4.2	4.5	4.7	5.0

Note.—Each figure in the table is the average actual rating given to those who estimated that that person would rate them a given way.

Examination of Table 7 shows that there was very little knowledge beforehand of how others in the group felt about them. This is of course to be expected. The groups were composed of people who were strangers to one another and, at the time these ratings and estimates were made, knew nothing about each other except their name, what part of the country they came from, and what they were studying. The Minneapolis data show no relation between estimates made and actual preference. The Leiden data do show a slight relationship. The average ratings tend to go up as the estimates increase. The differences for both conditions between the average ratings for estimates of 3 and estimates of 5 are significant beyond the .05 level. It is possible that in Holland, name, part of country, and field of study are more important items of information than they are in the United States. Such a difference, which it is plausible to assume, would account for the slight evidence of knowledge in the Leiden data before the discussion.

There are, of course, no differences between the compatible and incompatible conditions in either country. It would indeed be strange if there were differences since the actual manipulation which created the distinction between the conditions was performed after the initial ratings and initial estimates had already been made.

The same kind of analysis of the ratings and estimates made after the discussion will show

TABLE 8

ACCURACY OF ESTIMATION OF PREFERENCE AFTER THE DISCUSSION

Condition	Estimate of Others' Preference for Themselves				
	1 or 2	3	4	5	6 or 7
Compatible (M)	5.1	4.9	5.0	4.9	5.3
Incompatible (M)	4.5	4.9	5.1	5.0	5.7
Compatible (L)	4.6	4.3	4.5	4.9	5.3
Incompatible (L)	4.1	4.5	4.6	4.9	5.4

us whether or not *S*s learned something during the discussion about how others felt toward them. These data are presented in Table 8.

In Table 8 there is clear evidence that *S*s in the incompatible condition had significant knowledge, after the discussion, of how others in the group felt about them. This is true for both countries. The average rating received from those they think dislike them is significantly lower than the average ratings received from those they think will rate them 3, 4, or 5. These differences are significant at the .05 level or better in each country. Those that the incompatible *S*s think rate them 6 or 7 (like very much) actually do give them higher average ratings. In each country, these average ratings are significantly higher than any of the others ($p < .05$).

The evidence is not nearly as clear that *S*s in the compatible condition have knowledge of how others feel about them. In neither country is there any difference among the average ratings received from those who *S*s thought rated them anywhere from 1 (dislikes) through 5 (likes me moderately). Those that the compatibles think rated them 6 or 7, however, do give them somewhat higher ratings. In the Leiden data the differences between this average rating received from those they think will rate them 6 or 7 and any of the other average ratings are significant ($p < .02$). In the Minneapolis data the differences are smaller and only the difference between the average ratings from those who *S*s think will rate them 6 or 7 and those they think will rate them 4 or 5 is significant at the .05 level. But still, the data for both countries are extremely similar. The compatibles seem to have correct knowledge of how others feel about them only for those who they think like them very much. The incompatibles, on the other hand, also have correct knowledge about those they think dislike them.

The differences between the two conditions, while in the same direction for both countries, are larger for the Minneapolis data than for the Leiden data. In the Minneapolis data the difference between the compatible and incompatible conditions in the average rating that *S*s receive from those they think dislike them is significant at the .10 level of confidence; between those they think like them very much it is significant at the .06 level of confidence. The conjunction of both these differences is certainly highly significant. In the Leiden data the difference between the compatible and incompatible conditions in the average rating that they receive from those they think dislike them is significant at only the .20 level of confidence. No other differences between conditions approach significance. The fact that all the differences are in the same direction in both countries, however, lends strength to the assertion that the difference is a real one.

In comparing the data between the two countries one may note that the ratings made by the Minneapolis *S*s are consistently higher than the ratings made by the Leiden *S*s. Nothing can be made of this, however, since one cannot be certain that the translation of the rating scales from English into Dutch preserved intact the distances between points on the scale or the exact location of each point on the scale.

It will be recalled that in the introduction the hypothesis was advanced that *S*s in the incompatible condition, because of their instability and uncertainty about their relations with others in the group, would be more sensitive to the cues that others were giving and would know more accurately how others in the group felt about them. This part of the hypothesis has, in general, been confirmed by the data. It was also stated in the introduction that one would expect *S*s in the compatible condition, because of their greater certainty in the situation, to be more able to directly express their feelings toward others. If this hypothesis were true we would expect to find that *S*s knew the feelings of those in the compatible condition more accurately than they knew the feelings of those in the incompatible condition. This is not borne out by the data. There is no difference in accuracy of knowledge about the preferences of *S*s in the two conditions.

On the other hand, in Table 4 we observed that there was some difference between the two conditions in the nature of the cues concerning their preferences for others. Those in the compatible condition spoke more to those they disliked and less to those they liked than did *S*s in the incompatible condition. These data are still consistent with the notion that *S*s in the compatible condition

were more able to directly express their dislikes. However, since no difference was found in accuracy of knowledge about *S*s in the two conditions, it seems necessary to assume that other, equally good cues were made available by *S*s in the incompatible condition. At any rate, the evidence from the experiment gives no direct support to the second part of our initial hypothesis.

SUMMARY AND CONCLUSIONS

The major findings from the experiment which has been described may be summarized as follows:

1. If persons in a group feel that those members of the group whom they like best dislike each other, this tends to make them uncertain and unstable about their interpersonal relations in the group.

2. Both those who are stable and those who are unstable about their personal relations reveal something about how much they like others in how often they talk to them. They tend to talk least to those toward whom they feel indifferent.

3. There is some difference between the two conditions in how sensitive individuals are to the cues which others give off. Those who feel unstable about their personal relationships tend to have more accurate knowledge of how others feel about them. It is perhaps plausible to infer that those who are unstable about their personal relations watch more closely and are able to make better use of the available cues.

4. The results in both countries, Holland and the United States, are remarkably similar. The consistency of the results tends to strengthen confidence in their significance and also tends to show that these particular findings are independent of the cultural differences between these two countries.

REFERENCES

1. HEIDER, F. Social perception and phenomenal causality. *Psychol. Rev.*, 1944, **51**, 358–374.
2. HEIDER, F. Attitudes and cognitive organization. *J. Psychol.*, 1946, **21**, 107–112.
3. HEIDER, F., & SIMMEL, M. An experimental study of apparent behavior. *Amer. J. Psychol.*, 1944, **57**, 243–259.

Received June 10, 1953.

PROJECTION AND DISPLACEMENT: A CROSS-CULTURAL STUDY OF FOLK-TALE AGGRESSION[1]

GEORGE O. WRIGHT

Graduate School of Education, Harvard University[2]

THE STUDY of societies reveals many examples of the use of folk tales and rituals which are aggressive in content. In varying degrees this aesthetic aspect of culture seems to provide a device by which aggressive feelings can be displayed in fantasy without the disruptive effects which otherwise would result from their direct expression. This fantasy aggression, however, presumably has some correspondence with real-life aggression. In order to understand folk-tale aggression, therefore, we will attempt to discover the relationship between specific antecedent conditions and the specific outcomes with which they are associated in folk tales.

Part of the theoretical framework for the study of folk-tale aggression grows out of the approach-avoidance theory of Miller (1). This theory gives us a scheme for appraising the joint effects of approach and of avoidance in conflict situations. While the experimental evidence supporting the approach-avoidance theory originally came from work with animals, a reasonable elaboration will transform its application from psychophysical phenomena to those which are psychocultural.

One of the important elaborations of the approach-avoidance theory is the result of the work of Whiting (3, 4). Whiting's hypothesis asserts that the intensity of the displaced responses, as well as the cultural distance of displacement, is a function of anxiety. This hypothesis is confirmed in the doll-play investigations in which it was found that the display of large amounts of aggressive behavior in fantasy is connected with a high degree of anxiety arising from previous punishment for aggression (2).

Another elaboration of the approach-avoidance theory is the result of the work of Whiting and Child (4), in which the original theory as it applied to displacement is extended to projection. This work, which examined the relationship between aggression anxiety and the theories held by members of the society regarding the causes of disease, attested the validity of the theory and provided further evidence that such a theory, in fact, could be useful in understanding cultural phenomena.

The work of Whiting and Child did not independently test displacement and projection. This present study, however, does make this independent test by analyzing folk-tale behavior. The choice of folk tales for study was somewhat dictated by the fact that such an analysis could be carried out on the basis of classifying the behavior of the characters in the folk tales as either displacement (qualities of the objects of aggression) or projection (qualities of the agents of aggression).

PROCEDURE

The material used in this study has been developed from an analysis of folk tales from 33 societies, including selections from Oceania, Asia, Africa, North America, and South America. Since the child-training data of Whiting and Child (4) were to be used as one of the variables, folk tales had to be assembled for the societies that were among the 75 used in their study. The folk-tale literature of these 75 societies was examined, and preliminary selections on the basis of available material from 37 societies were made. At the outset it was decided not to include origin legends because of their lack of comparability with ordinary folk tales. For the same reason, those folk tales told exclusively in the third person with the ethnographer as narrator were excluded, since they were likely to be overlaid with the cultural bias of the ethnographer. In consequence of one or the other of these restrictions, the folk tales from four of the initial selections were discarded. These societies were the Ainu, Arapesh, Azande, and the Maori.

In order to provide sufficient data for analysis, twelve folk tales were selected from each society. Moreover, in order to eliminate the inclusion of fragmentary tales, each of the folk tales selected was required to be a complete episode. An episode was defined as a sequence of action having a definite beginning and proceeding to some definite action climax.

In order to carry out the study, a scheme permitting quantitative analysis of the folk tales was needed. In addition, it was necessary that the scheme comprise ob-

[1] This material is part of a doctoral dissertation submitted by the author to the Graduate School of Education, Harvard University, 1952. The author wishes to express his appreciation to Dr. John W. M. Whiting for his guidance in carrying out the research on which this paper is based.

[2] Now at the Wright Air Development Center, U. S. Air Force.

jective scoring categories so that the subjective estimates of the investigator would be minimized. These requirements are met in the scoring scheme used by the doll-play investigators. Their scheme was modified for use with folk tales.

Central to the scoring system is the concept of *interaction units*. As the term implies, an interaction unit is a complete behavior sequence between two persons. In order to be complete, the sequence must contain at least two action elements. First, a person must initiate an action toward some other person. Second, this initiated action must be responded to by the other person. For example, a mother initiates action by calling her child to lunch. The child responds to this action by stopping his play and coming into the house. The first element we can call *initiation*; the second, *response*. The person starting the action we can call the *initiator*; the person responding to the action we can call *respondent*.

Now, since many of the actions are likely to be more than simple action-reaction, we have decided on a convention for complex actions. In our example of the mother calling the child to lunch, how will the action be scored if the child were to say, "I'm busy, mother, and can't come in now"? What is the action of the mother to be called if she in turn responds to the reply of the child by a threat of spanking? The convention adopted is to score the calling of the mother and the reply of the child as one interaction unit, with mother as initiator and child as respondent. In turn, the reply of the child and the threat of the mother are scored as the second interaction unit, with the child as initiator and the mother as respondent. If the threat of punishment causes the child to come into the house, this is scored as the third interaction unit, with mother as initiator and child as respondent. Allowing a reaction to be both a response and an initiation solves our scoring dilemma.

When the interaction unit is used as the vital measure in the scoring system, it is possible to build additional classifying measures around this unit. For example, it is possible to classify the status of the initiator, the status of the respondent, the nature of the initiation, and the nature of the response. It is possible, moreover, to make comparisons from folk tale to folk tale by the use of the quantitative measures.

The interaction sequences were scored as aggressive or nonaggressive; each aggressive act was scored for intensity (ranging from a score of 1 for nonaccomplishment of aggressive intent to a score of 5 for killing); and finally, each sequence was classified as to the agent and object of the act.

In order to test the reliability of the scoring system, four students at the Wilberforce State College (Ohio) were asked to score 36 folk-tale episodes selected at random from the collection. These students were trained in the use of the system for five hours prior to the scoring. An index of reliability was computed as follows: the sum of the agreements for a pair of the student scorers was divided by the total number of items scored by them; the scores of the writer and of the students were randomly paired in all possible combinations for each of the scoring categories; the resultant index of reliability was 0.83, with 1.0 representing perfect agreement.

The scoring of the folk tales provides a comprehensive set of data on the display of aggressive feelings in folk tales. This is our consequent variable. Data also were needed for the child-training practices, our antecedent variable. For this antecedent variable, the child-training data of the study of Whiting and Child (4) were used. These data, in the form of ratings on seven-point scales, were available for 26 features of child-training practices. The details of the rating scale for the practice that is relevant to the present study are as follows (the numbers indicate the ends of the rating scale):

Severity of Aggression Training

7—Child severely punished for any form of aggression; fighting, disrespect, or disobedience not permitted.

1—All but the most severe forms of aggression permitted; fighting encouraged.

In analyzing the data, the societies were divided into two groups, high and low, on the basis of the ratings for the child-training practice. The data were then analyzed for significance of difference between means.

THE EFFECT OF PUNISHMENT ANXIETY ON DISPLACEMENT

The first hypothesis to be tested is that objects of displaced aggression are chosen farther and farther out on the generalization continuum as a function of the amount of aggression anxiety.

This hypothesis is derived from the approach-avoidance theory, which considers the joint habit strength of aggression and aggression anxiety to be the algebraic difference between the two tendencies. As the degree of anxiety is increased, this anxiety function shifts the point of maximum response toward objects less and less like the original object.

The approach-avoidance theory specifies *where* on the generalization continuum the maximum response is likely to occur, but Whiting's theory specifies also *how strong* this maximum response is likely to be. This theory asserts that anxiety is frustrating, that is to say, it interferes with the initial response, and that this interference is itself drive-producing. Consequently, the joint effective habit potential is a function of the product of the joint habit strength (i.e., of the difference between the aggression and the aggression anxiety) and the strength of frustration as indicated by aggression anxiety.

In order to translate this hypothesis into terms that can be tested with the folk-tale data, the punishment of children for aggression will be called aggression anxiety; the Hero will be called the point of original response; the similarity dimension will be considered as beginning at the Hero and extending through

friends, relatives, and acquaintances to strangers at the least similar position; and finally, the frequency of aggressive acts involving a given category along the dimension of similarity will be called strength of response. Further, we will consider our hypothesis supported if the following conditions are found:

1. In the folk tales of a society in which the child-training practice is low severity of punishment for aggression, the frequency with which friends of the Hero are used as objects of aggression should be higher than it is in the folk tales from societies in which the practice is high severity of punishment for aggression. This condition should be found because our theory predicts that the lower the anxiety, the more like the original object will be the displaced objects.

2. Conversely, in the folk tales of societies in which the child-training practice is high severity of punishment for aggression, the frequency of the use of strangers as objects of aggression should be higher than in those of societies in which the practice is low severity of punishment for aggression. This condition should be found because our theory predicts that the higher the anxiety, the less like the original object will be the displaced objects.

The data used for this analysis were the antecedent child-training measures of severity of punishment for aggression versus the frequency of use of the friends of the Hero as objects of aggression. The scores of frequency of use of friends as objects of aggression were obtained by counting the number of times the friend is used as the object of aggression in the folk tales, and dividing this frequency by the total aggressive acts.

Analysis of the data indicates that for the group of societies in which aggression anxiety is low (low severity of punishment for aggression), the mean frequency of the use of friends of the Hero as the object of aggression is 13.3, whereas the mean of the group of societies in which aggression anxiety is high (high severity of punishment for aggression) is 8.67. The difference is significant at the .005 level (obtained $t = 3.8$).[3] The results of this and the following tests are shown in Table 1.

The use of the measures for the friends of the Hero as the object of aggression instead of the Hero himself is a question that naturally arises. In applying the test of the hypothesis, it was felt that the Hero as the object of aggression actually represented self-aggression, and that this fact might introduce this effect as an uncontrolled variable. Consequently, the objects considered closest to the Hero, namely, his friends, were chosen.

In order, however, to provide information that might be useful in refining the hypothesis, the analysis of the data on the Hero was carried out. In this analysis, the means of the frequency of use of the Hero as object of aggression were compared for the high aggression-anxiety group of societies and the low aggression-anxiety group of societies. The difference is not significant—that is to say, the difference in the means which we observed are likely to have arisen by chance. The obtained t is 0.47; the significance level is approximately .30.

The test of the converse condition, that of comparing the frequency of the use of strangers as the object of aggression under the conditions of high aggression anxiety and of low aggression anxiety, discloses that the mean of the frequency of use of strangers as the object of aggression for societies in which aggression anxiety is high is higher than in societies in which aggression anxiety is low. The means are respectively 31.66 and 26.58. The difference between the means is significant at approximately the .02 level (obtained $t = 2.21$). Thus, in sum, the societies in which children are severely punished for aggression,

TABLE 1

SUMMARY OF FINDINGS

Mean Frequencies of Aggression	Aggression Anxiety		t	p
	High $N = 12$	Low $N = 14$		
Agent of aggression				
Hero as agent	22.3	23.0	0.21	.25
Friend as agent	15.33	18.54	1.18	.13
Stranger as agent	30.5	25.43	2.46	.01
Object of aggression				
Hero as object	27.33	29.43	0.47	.30
Friend as object	8.67	13.3	3.8	.005
Stranger as object	31.66	26.58	2.21	.015
Intensity of aggression	19.58	15.7	2.02	.035
Hero triumphant	2.14	4.85	2.86	.005

Note.—This table shows mean scores for frequencies of various categories of agents and objects of aggression, for intensity of aggression, and for fate of the hero in the folk tales. (Societies are dichotomized on the basis of two degrees of aggression anxiety, high and low.)

[3] In all of the results reported, only one tail of the t distribution is used, since all of the results are in the predicted direction.

as compared to the mild societies, represent friends of the Hero *less* frequently and strangers *more* frequently, as would be expected from the displacement hypothesis.

THE EFFECT OF PUNISHMENT ANXIETY ON PROJECTION

The second hypothesis to be tested is that the agents of projected aggression are chosen farther and farther out on the generalization continuum as a function of the amount of aggression anxiety.

In order for our theory to be consistent, the findings for displacement should have their counterpart in projection. The child-training measures of punishment aggression, used as indices of aggression anxiety, should be related to projection in such a way that the folk tales of societies in which punishment for aggression is severe should display more aggression performance that uses the stranger as agent, since the tendency to express aggression under these conditions should be projected far out on the generalization continuum. Stated another way, high aggression anxiety should be associated with a more frequent use of stranger as aggressive agent.

The folk-tale data translated into terms of our theory are analyzed, using severity of punishment for aggression as the degree of anxiety about aggression, and the frequency of the use of various categories as agents of aggression; these were considered an index of the strength of the joint effective habit of aggression and aggression anxiety. We will consider our hypothesis as supported if the following conditions are found:

1. In the folk tales of societies in which the child-training practice is low severity of punishment for aggression, the frequency of use of the Hero as agent of aggression should be higher than in those of societies in which the practice is high severity of punishment for aggression. This condition should be found because our theory states that the lower the anxiety, the more like the original object will be the projected agents. This condition should also be found for the use of friends of the Hero as agents of aggression.

2. In the folk tales of societies in which the child-training practice is high severity of punishment for aggression, the frequency of use of strangers as agents of aggression should be higher than in folk tales from societies that practice low severity of punishment for aggression. This condition should be found because our theory states that the higher the anxiety, the less like the original objects will be the projected agents.

The data used for this analysis were the antecedent child-training measures of severity of punishment for aggression versus the frequency of the use of the Hero as agent of aggression. The scores of frequency of use of the Hero as agent of aggression were obtained by counting the times the Hero was used as agent of aggression in the folk tales, and dividing this by the total aggressive acts.

Analysis of the data indicates that the mean frequency of use of Hero as agent of aggression in the high aggression-anxiety societies is 22.3, whereas it is 23.0 in the low aggression-anxiety societies. The difference between these means, though in the predicted direction, is not significant (p is approximately .25; obtained $t = 0.21$).

Analysis of the data for the use of friends of the Hero as agent of aggression, using a procedure similar to that for the Hero, indicates that the difference between the means of frequency of use of friends as agents of aggression is also in the predicted direction, but is not significant (p is approximately .13; obtained $t = 1.18$).

In the converse condition, namely, high aggression anxiety accompanied by the more frequent use of the stranger as agent of aggression, analysis of the data indicates that for the group of societies in which aggression anxiety is high, the mean frequency of use of stranger as agent of aggression is higher than that of the group of societies in which the aggression anxiety is low. The mean frequencies are, respectively, 30.5 and 25.43. The difference between the means is in the predicted direction and is significant at the .01 level (obtained $t = 2.46$).

The findings discussed here tend to support the theorems and to suggest that the individual may, in fact, use psychologically harmless ways to express his aggression in the presence of aggression anxiety. It is suggested, moreover, that the individual apparently chooses harmless agents and objects of aggression on the basis of their categorical differences from the agents and objects associated with the initial anxiety. The findings also suggest that two qualifications are essential in the inter-

pretation of these data. Both relate to the discrete placement of the cultural categories along the dimension of similarity.

The first qualification is that the present theory, which relates anxiety to projection and displacement, is not refined enough to predict the characteristics and the behavior of the Hero in folk tales as a function of the anxiety of the society. The explanation proposed is that the position which the Hero occupies on the generalization continuum is theoretically influenced by the transition between real life and fantasy and that more things must be known about the influence of complex psychological factors in this transition area before our predictions will gain accuracy.

The other qualification is that a precise placement of the intermediate categories of relatives and acquaintances cannot be made with the use of our presently developed theory. The explanation of this condition probably lies in the complexity of meaning which various societies attach to cultural categories. It is possible, for example, that in one society the similarity of relatives to strangers might be perceived to be greater than between acquaintances and strangers, while in another society the opposite might be true.

THE EFFECT OF AGGRESSION ANXIETY ON INTENSITY OF AGGRESSION

Another aspect of the relationship of aggression anxiety to fantasy aggression is of interest, namely, the apparent influence of the strength of anxiety on the intensity of aggression in folk tales. Our theory suggests that the aggression anxiety will diminish at a more rapid rate than the intensity of the aggression as displacements and projections are made farther and farther out on the generalization continuum. We should be able to foretell from this theory that fantasy, since it represents situations in many ways dissimilar from real life, should provide a vehicle for the expression of aggression, and should do so with an intensity which is directly related to the amount of real-life punishment for aggression.

An examination of this theory is made in the third hypothesis to be tested, namely, that punishment for aggression increases the intensity of aggression in situations which are distinctly dissimilar from those in which the punishment occurs.

This hypothesis, translated into terms which are applicable to the folk-tale data, states that the punishment of aggression leads to aggression anxiety. This anxiety reduces the display of aggressive behavior toward objects similar to those associated with the initial punishment. As we have seen, in such situations displacement and projection of aggression are likely to occur. Moreover, the strength of the joint effective habit of the aggression and of the aggression avoidance is thought to be the product of the difference between the two tendencies and the degree of avoidance. We should expect, therefore, that the greater the anxiety, the greater the intensity of the fantasy aggression.

Our hypothesis will be supported if it is found that the folk tales of societies in which high aggression anxiety is generated by severe punishment for aggression show a greater intensity of aggression than exists in the folk tales of societies in which low aggression anxiety is generated.

The data used for this analysis were the same child-training measures of severity of punishment for aggression used in the previous analysis related to the intensity of the folk-tale aggression. The scores for intensity were obtained by taking the sum of the intensity measures for each act of aggression in the folk tales of a given society and dividing this sum by the total acts of aggression in the folk tales.

An analysis of the data indicates that the mean of the intensity of aggression for the group of societies practicing high severity of punishment for aggression is 19.58, whereas the mean for the group practicing low severity of punishment for aggression is 15.7. The difference between the means is in the predicted direction and is significant at approximately the .03 level (obtained $t = 2.02$).

THE EFFECT OF AGGRESSION ANXIETY ON THE FATE OF THE HERO

There is one additional analysis by which information may be disclosed about the fantasy structuring of the aggression in folk tales, namely, the relationship between aggression anxiety and the fate of the Hero in the folk tales. Actually, such an analysis provides another test of the kind and amount of projection and displacement which develop as the consequent variable of aggression anxiety. The fate of the Hero, moreover, is likely to be a more sensitive measure of the expression of

fantasy aggression than those involving secondary characters.

An examination of this theoretical position is done in the fourth hypothesis to be tested, namely, that the Hero should triumph less often in folk tales of societies in which the child-training practices include severe punishment for aggression. This hypothesis is a restatement of the second hypothesis that high aggression anxiety results in projection far out from the Hero on the generalization continuum. This theorem is built on the assumption that the behavior of the Hero is understandable as a projection of the individual. Consequently, in cases of high aggression anxiety, it is to be expected that having the Hero triumphant is likely to be accompanied by large amounts of anxiety, so that triumphant characters of the folk tales are likely to be less like the Hero. In short, a small degree of triumphant-Hero aggression should accompany high aggression anxiety.

The data used for this analysis were the child-training data of severity of punishment for aggression versus the frequency of Hero triumphant in the outcomes of the episodes.

An analysis of the data for the fate of the Hero shows that the Hero is triumphant less frequently in the high aggression anxiety group of societies. The means are 2.14 and 4.85. The difference between the means is significant at the .005 level (obtained $t = 2.86$).

This finding is in accord with the earlier results in which it was seen that a comparable relationship was true, namely, that of the more frequent use of the stranger as agent of aggression in folk tales of societies in which aggression anxiety is high. In both cases, the theoretical assumptions underlying the functioning of the phenomenon of generalization in conflict are found to be valid.

SUMMARY

1. The analysis of the folk-tale data supports the essential features of the theory which was developed to supply insight into the consequences and outcomes of action and behavior in folk tales. The approach-avoidance theory of Miller, restated in terms of Whiting's modification of it for the influence of conflict-produced drives, provides a coherent basis for the explanation of the phenomena of projection and displacement in folk tales. Admittedly, the predictive powers of such a theory are not perfect, but the results obtained for the tests of the hypotheses are all in the predicted direction.

2. In general, it is found that high punishment anxiety, when it is connected with the expression of aggression, leads to the kind of folk-tale behavior in which the objects and agents of aggression are far out on the generalization continuum and are least like the Hero. In addition, the Hero in these folk tales is not likely to be triumphant in the outcomes.

3. In general, it is found that low punishment anxiety, when it is connected with the expression of aggression, leads to the kind of folk-tale behavior in which the objects and agents of aggression are chosen close to the Hero on the generalization continuum. In addition, the Hero is likely to be triumphant in the outcomes.

4. High punishment anxiety, when it is connected with the expression of aggression, leads to the display of a more intense kind of aggressive behavior in folk tales than is true for low punishment anxiety.

5. Finally, the study demonstrates that folk tales as fantasy material are fruitful for analysis. Enough consistency of trends has been found in the present study to suggest that the relationships of child-training practices and the expression of aggression in folk tales are real and not accidental or spurious. This consistency and its value are somewhat enhanced by the fact that the variables of child-training practices used in this analysis are data from an independent investigation which were collected and analyzed by workers who were not directly involved in the scoring and analysis of the present folk-tale study.

REFERENCES

1. DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
2. HOLLENBERG, ELEANOR, & SPERRY, MARGARET. Some antecedents of aggression and effects of frustration in doll play. *Personality*, 1951, **1**, 32–43.
3. SEARS, R. R. A theoretical framework for personality and social behavior. *Amer. Psychologist*, 1951, **6**, 476–483.
4. WHITING, J. W. M., & CHILD, I. L. *Child training and personality*. New Haven: Yale Univer. Press, 1953.

Received June 22, 1953.

A MOVEMENT INTERPRETATION OF THREAT[1]

ALVIN W. LANDFIELD

Purdue University

THE PRESENT investigation of threat was carried out within a theoretical framework recently advanced by Kelly (2). The independent and dependent variables, the relationship between them, and the methodology employed in testing the experimental hypotheses are derivable from or are consistent with the following theoretical statements of the psychology of personal constructs.

A person's psychological processes evolve toward what he construes to be an optimal anticipation of events. Definitions: Processes—. . . the subject of psychology is assumed at the outset to be a process. Evolve—. . . the patterns of psychological processes undergo progressive change in themselves as they function. Toward—Directionality of processes is built into the fundamental postulate. What he construes—The directionality of processes is assumed to be an essential function of prior outcomes of those same processes. The person himself sets the objectives and he does it, just as the psychologist must, by construing, by attributing significance, by framing events with meaning at some minimal level. Anticipation—The direction of mental processes is established as prediction. This is an extrapolation to the domain of the individual of the commonly accepted notion regarding the aim of science being prediction and control. Optimal—. . . some sort of reconciliation of potentially contradictory anticipations must be assumed as taking place in the individual . . . (2, pp. 29–31).

Threat is the awareness of imminent comprehensive change in oneself. Definitions: Comprehensive change—Employed in the context of threat, it refers to a modification in the directionality of evolving psychological processes. Restated, it refers to the change in the direction of change. Not all psychological change implies a modification in directionality.

Self-movement, the independent variable, was chosen as one measure of an individual's evolving psychological processes. Self-movement refers to the direction of change in self-conceptualization, inferred from the different ways an individual describes himself (*a*) as he was in the past, (*b*) as he is now, (*c*) as he would like to be in the future. The perception of threatening and nonthreatening acquaintances, in terms of roles and expectancies, constituted the dependent variable. Definitions: Role—the way an individual rates another person. Expectancy—the way an individual thinks another person might rate him. Threatening acquaintance—one whom an individual wishes to avoid or whose behavior he would like to modify greatly. Nonthreatening acquaintance—one with whom an individual wishes to have contact and whose behavior he would not care to modify greatly. The description of acquaintances in terms of roles and expectancies is consistent with our theoretical framework. Roles and expectancies are common ways of anticipating or predicting other people.

Since we have employed self-movement as one measure of evolving psychological processes, threat may be conceptualized as an imminent change in the directionality of self-movement. Therefore, according to our theory, the threatening acquaintance described in terms of roles and expectancies is threatening because interaction with him might change the directionality of self-movement. Now if threat is related to the change in direction of self-movement, what kinds of roles and expectancies might lead to such a change? Our hypothetical answer to this question may be stated as follows: The threatening other person will be perceived by an individual as being more like himself as he was in the past, or is now but no longer wants to be, than the nonthreatening person. Also, the threatening other person will be seen by an individual as expecting him to be more like he used to be, or is now but no longer wants to be, than the nonthreatening person. The latter formulation, known as the expectancy hypothesis, may be readily illustrated. The adolescent is threatened by his parents, not merely because he feels he is sometimes perceived as a child, but, because the parents seem to perceive him as a child, he might begin acting like a child. Because the parental expectancy is so plausible, the directionality of his self-movement toward maturity could easily be reversed.

If we can show that a relationship exists

[1] This paper is based upon a doctoral dissertation completed in 1951 at The Ohio State University under the direction of Dr. George A. Kelly.

between the axes along which an individual perceives himself changing and his perceptions of threatening people, then we may conclude that our findings tend to support the more comprehensive formulation of threat within the psychology of personal constructs.

METHOD

A social concept test (1) and rating scales derived from this social concept test were employed as research tools. In the social concept technique, a subject (*S*) writes the names of certain specified persons he knows, 30 in all, on separate cards. These cards, a name on each one, are then dealt to *S* three at a time, in certain combinations. As three cards are dealt, *S* is asked, "In what outstanding way are two of these persons similar and different from the third?" From *S*'s responses to this question, rating scales may be created, scales which are derived from *S*'s descriptions of these people and are expressed in his own terminology. For example, *S* may state that two of the persons are "gregarious" and the other person is "not gregarious." The terms "gregarious" and "not gregarious" are placed at either end of a nine-point rating scale. It should be emphasized at this point that *S*s do not always give opposite concepts in expressing a similarity and difference between three people. To insure opposite concepts (i. e., opposite to *S*), the examiner, when in doubt, simply asks whether or not the two descriptions are opposites. If the descriptions are not in opposition for *S*, he is asked to state their opposites.

The *S* is then asked to place two or three persons with whom he is acquainted at points on each scale, depending upon whether he sees the scale as dichotomous or dimensional. Whenever *S* feels he can use only the poles of his scale, it is considered to be dichotomous and only the poles of the scale are used for ratings. Two referent persons then are placed on the scale, one at each end of the nine-point scale, and scale points 2 through 8 are crossed out. Whenever *S* perceives people distributed across his scale, it is considered dimensional, and all points of the scale may be used for the ratings. Three referent persons then are placed on the scale, one at each end of the nine-point scale and one in the middle. In other words, scale points are represented by people. (See Fig. 1.)

Both exploratory and validation designs were employed. Fifteen adult *S*s were used in the exploratory phase of the study. Fifty-four volunteer undergraduate male and female *S*s, aged 18 through 24, drawn from economics and psychology classes at The Ohio State University, were used in the validation study.

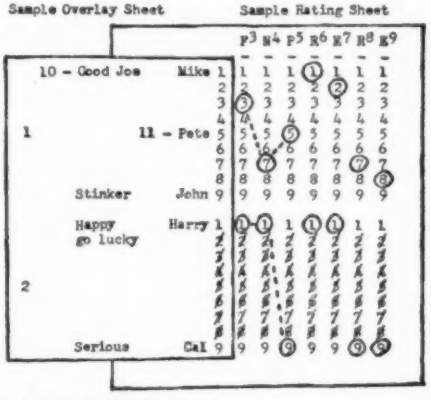
The social concept test was administered individually during one experimental session to each *S*. The 30 names used in this technique were specified by the experimenter (*E*) in order to insure adequate sampling of persons with whom an *S* is acquainted. For example, the following people were specified: teacher you like, mother, father, sister nearest your age or person like a a sister to you, person with whom you would not care to go on a vacation; etc.

The rating scales derived from *S*'s responses to the concept test were administered individually during a second experimental session, at which time *S* was asked to: (*a*) rate himself on his own scales as he would like to be six years from now, if it were possible for him to be anything he wished; (*b*) rate himself as he feels he is now; (*c*) rate himself as he feels he was at some period during the past six years when he was most different from the way he is now; (*d*) think of a threatening and a nonthreatening person and rate each of them in terms of *roles* and *expectancies*. Threatening and nonthreatening persons were of the same sex as *S*.

As can be seen in Fig. 1, all ratings were done on a rating sheet and the rating scale labels were placed on a separate overlay sheet. The *S* places the right-hand edge of the first overlay sheet to the left of the first column, and using the name designations as reference points, rates the first six scales by encircling the proper numbers in the first column on the rating sheet. The *S* then takes the second overlay sheet with a different six scales and the second rating sheet and continues rating in this first column. After all scales have been rated in the first column, *S* takes the first overlay sheet again, moves it one column to the right to the second column, thereby covering over the first column, and proceeds as before. This procedure is continued until *S* has completed all columns.

SELF-MOVEMENT

Employing the *role* and *expectancy* descriptions of the threatening and nonthreatening persons as dependent variables, we attempted to predict these descriptions from the directionality of self-movements, inferred from the change in self-description from past, to present,



1 A dimensional scale
2 A dichotomous scale
3 Future ideal rating column
4 Present rating column
5 Past rating column
6 Role rating column, nonthreatening person
7 Expectancy rating column, nonthreatening person
8 Role rating column, threatening person
9 Expectancy rating column, threatening person
10 A concept description
11 Acquaintance placed on scale as rating reference point

FIG. 1. RATING FORMS

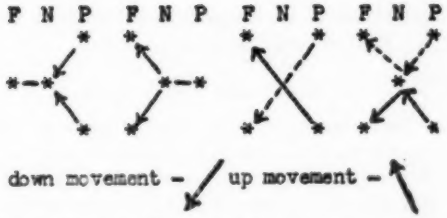


FIG. 2. TYPICAL MOVEMENTS INFERRED FROM PAST, PRESENT, AND FUTURE SELF-RATINGS

to future. A self-movement is inferred to have occurred within a single concept scale when there is a numerical difference between the past and present or present and future ideal self-ratings. It should be noted that the movement shown in Fig. 1 is from right to left. Our use of the overlay sheet necessitated this arrangement.

Self-movement may occur between the past and the present, directionality up or down; between the present and the future ideal, directionality up or down; and between the past and future ideal, directionality up, down or up and down (i.e., the movement between both past and present and present and future ideal may be up. It also may be down. Then again, one movement may be up and the other down. The latter is called bidirectional movement). (See Fig. 2.)

The independent variable more specifically was a unidirectional self-movement within a concept scale. Where the movement from past to present to future ideal is unidirectional, prediction simply is made from the directionality of this unidirectional movement. (See Fig. 3.)

Where the movement from past to present to future ideal is bidirectional, prediction is made from the longest unidirectional move-

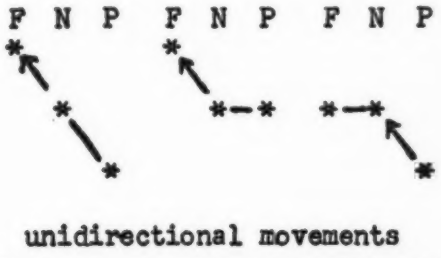


FIG. 3. UNIDIRECTIONAL MOVEMENTS INFERRED FROM PAST, PRESENT, AND FUTURE SELF-RATINGS

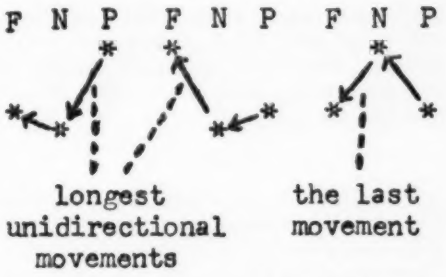


FIG. 4. BIDIRECTIONAL MOVEMENTS INFERRED FROM PAST, PRESENT, AND FUTURE SELF-RATINGS

ment. (The frequency with which bidirectional movement occurred was small. Movement usually occurred only from past to present, only from present to future, or in the same direction from past to present to future.) (See Fig. 4.)

In the few cases in which the two opposite movements are the same in length, prediction is made from the last movement (Fig. 4).

As we previously stated, the threatening and nonthreatening other persons were described in terms of *roles* and *expectancies*. These descriptions, expressed as numerical ratings, were predicted from the direction of a unidirectional self-movement, relative to each concept scale. In other words, a set of predictions was made for every concept scale within an *S*'s scale repertoire in which movement occurred.

EXPERIMENTAL HYPOTHESES

Role hypothesis. The *S*s will attribute the *role* rating nearest the end of the concept scale toward which the unidirectional self-movement is pointing *more often* to the nonthreatening person than to the threatening person.

Expectancy hypothesis. The *S*s will attribute the *expectancy* rating nearest the end of the concept scale toward which the unidirectional self-movement is pointing *more often* to the nonthreatening person than to the threatening person.

The following example may clarify these hypotheses. An *S* has created 15 concept scales within which self-movement may be inferred. If the direction of movement successfully predicts the *role* ratings in ten scales and fails to predict the *role* ratings in five scales, it may be inferred that successful

predictions exceed unsuccessful predictions, and accordingly, the *S* conforms to the hypothesis. However, a significant number of *S*s must follow the same pattern before the role hypothesis is supported.

RESULTS

Forty-eight *S*s attributed the *role* rating nearest the movement-designated end of the scale more often to the nonthreatening person. Three subjects attributed the *role* rating more often to the threatening person. In three cases successful predictions equaled unsuccessful predictions. These latter cases were handled statistically by arbitrarily placing one and one-half cases with the successful prediction group and one and one-half cases with the unsuccessful prediction group. Chi square, with a 50:50 theoretical frequency and 1 degree of freedom, was employed in the analysis. The statistical results ($\chi^2 = 37.5$) allow the null hypothesis to be rejected at better than the .001 level of confidence.

Forty-nine *S*s attributed the *expectancy* rating nearest the movement-designated end of the scale more often to the nonthreatening person. One *S* attributed the *expectancy* rating more often to the threatening person. In four cases, successful predictions equaled unsuccessful predictions. These latter cases were handled statistically by arbitrarily placing two cases with the successful prediction group and two cases with the unsuccessful prediction group. Chi square again was employed. The statistical results ($\chi^2 = 42.7$) allow the null hypothesis to be rejected at better than the .001 level of confidence.

SUMMARY AND CONCLUSIONS

The results of this study support the hypothesis that a highly significant relationship exists between the axes along which an individual perceives himself changing and his perceptions of threatening people. More explicitly stated, the threatening person is perceived by an individual as being more like himself as he was in the past, or is now but no longer wants to be, than the nonthreatening person. Also, the threatening person is perceived by an individual as expecting him to be more like he used to be, or is now but no longer wants to be, than the nonthreatening person. Within our theoretical framework, these findings may be interpreted further to mean that the threatening person is placed by the threatened individual in the unique position of being able to reverse the individual's direction of movement, to interrupt his movement, or at least to place him under additional pressure to resist such modification of his movement.

REFERENCES

1. HUNT, D. E. Studies in role concept repertory: conceptual consistency. Unpublished master's thesis, Ohio State Univer., 1951.
2. KELLY, G. A. The psychological construction of life. Unpublished manuscript, Ohio State Univer., 1951.

Received June 22, 1953.

THE COLOR-FORM RESPONSES OF NORMAL, PSYCHOTIC, AND NEUROTIC SUBJECTS

J. D. KEEHN[1]

Maudsley Hospital, London

INVESTIGATIONS into the color-form attitudes of various special and clinical groups have been reported by several writers since Kuhlmann's (17) study of defectives in 1904. In addition to the large number of studies of normal adults and children (see Lindberg [18], Eysenck [7, 8], for reviews), there have been investigations of the color-form reactions of defectives (2), organics (11), spastic children (6), epileptics (1), psychotics (3, 4, 9, 15, 18), and neurotic patients (5, 7, 19).

Altogether some seven different types of color-form tests have been employed by these workers. Time of sorting by color or form was used by Kuhlmann (17) and by Eysenck (9), Kibler used differential memory of tachistoscopically exposed colored shapes, and Braat (3) instructed his subjects (*S*s) to match a given figure with a set of figures shown tachistoscopically. Grouping of symbols by color or shape formed the basis of tests employed by Goldstein and Scheerer (11), Cotton (6) and, in a modified form, by Eysenck (7). A test involving the ranking first of a set of colors, then a set of shapes, and third a selection of colored shapes—such that the most preferred shape appeared in the least preferred color—was given to a group of neurotics by Eysenck (7), and sorting tests were devised and administered by Lindberg (18), Clarke (5), and Eysenck (9). Finally the Ring test, designed and described by Lindberg (18), was used by Lindberg (18, 19), Campailla (4), Eysenck (7), and Clarke (5).

It was not until Eysenck (7) reported that the average intercorrelation between the three tests used by him was only .24 that the suggestion that the various color-form tests might not all be measuring the same thing was put forward. Thus Eysenck speculated that either the tests used by him were all measuring fairly specific attitudes or that if a general factor of color-form attitude existed, his particular tests might be rather poor measures of it.

Further doubts about the comparability of color-form tests arise from a study of the results obtained from the application of such tests to psychotic *S*s. Thus while schizophrenics appear able to abstract color from form or form from color more easily than manic-depressives in the tachistoscopic memory test (15), no significant differences between the attitudes to color or form of the two groups were found on the sorting test employed by Eysenck (9). Hence the tests seemed more likely to be measuring the ability to dissociate than to be indicative of attitude to color and form when used with these groups. However, this interpretation has not been extended to normal subjects, for Kretschmer and his associates (8) have used response to color and form as a distinguishing feature between schizothymic and cyclothymic personality types which are posited to be the normal prototypes of the more extreme psychoses.

In order to clarify the situation somewhat, a factor analysis of a number of color-form tests was carried out by Keehn (13, 14). A factor of color-form attitude was identified, but although this pointed to the comparability of the earlier tests in terms of method, the correlations between the tests were low enough to show that results obtained from the use of one test were of little value in predicting response to others. Hence it was felt (14) that a battery of tests based on the results of the factor analysis would give more valid and reliable results than those obtainable from the use of one test alone.

THE TEST BATTERY

From a large battery of color-form tests (13), nine were selected for special analysis (14), and it is these tests which constitute the present battery. The number beside each test refers to its serial position in the larger battery (13).

2. *Tossing.* A disc was tossed like a coin and *S* could respond to color or shape in guessing which side would fall uppermost.

4. *Memory: paired associates.* Four capital letters had to be associated with four colored shapes. Colors

[1] The author wishes to express his gratitude to Dr. H. J. Eysenck at whose suggestion and with whose advice and encouragement this study was carried out. This work was done while the author was a London University Postgraduate Scholar at the Institute of Psychiatry.

and shapes were later interchanged to see which category had been associated with the letters.

8. *Grouping.* Twelve squares set out on a background could be grouped either by color or by disposition.

10. *Odd one out.* Of nine figures on a card one was a lone shape and one a lone color. The subject was asked to pick out the odd figure.

12. *Stepping stones.* The subject had to pick his way across a set of colored shapes from one point to another. Two correct paths were available, one where all the shapes were the same, the other where the colors were the same.

13. *Common elements.* Three similar figures had to be chosen from eight. "Similarity" could be based on color or shape.

19. *Weigl sorting.* Nine wooden blocks could be sorted by color or shape.

22. *Card sorting.* A pack of cards was to be sorted into four compartments. All but four of the cards could be sorted in only one way; the remaining four could be placed either by color or by shape.

24*b*. *Tachistoscopic matching.* A given figure had to be selected from eight figures exposed tachistoscopically. Selection could be by color or form. No figure exactly matched the original.

The factorial analysis of these tests yielded two factors, the test saturations being reproduced in Table 1. The first factor was identified as that of attitude toward color or form. The obtained first factor saturations were used as though they were correlations of the tests with the factor (21), and regression weights for maximum predictability of the color-form factor by the battery were calculated according to the Ledermann method as set out in Thomson (21). The resulting weights are shown in Table 2. The ensuing multiple correlation of the battery with the factor was found to be .93 thus showing the battery, weighted in this fashion, to be a highly valid measure of the factor.

In order to obtain an estimate of the reliability of the battery, use was made of the generalized Spearman-Brown prophecy formula $r_T = \frac{Kr}{1 + (K\text{-}1)r}$ as set out in Gulliksen (12), where r_T is the estimated reliability of a total test whose inter-unit correlation is r and where the unit test is increased in length K times. This formula rests on the assumptions that all the units are psychologicaliy similar and that the correlations between them are all equal. In this case each individual test was taken as a unit and, as all nine tests have been shown to measure along a common dimension, the assumption of psychological similarity is not unreasonable. However, the correlations between the items varied between −.18 and +.44 with a mean of +.21, so that although the average of these correlations was used in the formula, the obtained reliability estimate of .71 must not be regarded as anything more than a general indication of the reliability of the battery.

TABLE 1

FACTOR SATURATION

Test	2	4	8	10	12	13	19	22	24*b*
F I	.33	.50	.38	.60	.57	.54	.52	.61	.26
F II	−.46	−.43	.45	.21	.14	.08	.44	−.24	−.17

TABLE 2

REGRESSION WEIGHTS

Test	2	4	8	10	12	13	19	22	24b
Weight	.20	.30	.06	.21	.19	.18	.15	.31	.09

TABLE 3

PERCENTAGE OF THE VARIOUS GROUPS GIVING THE COLOR RESPONSE TO THE INDIVIDUAL TESTS

Test	Percentage of *Ss* Giving the Color Response				Test	Percentage of *Ss* Giving the Color Response			
	Normals	Neurotics	Schizophrenics	Manic-depressives		Normals	Neurotics	Schizophrenics	Manic-depressives
2	61.5	48	64	60	16	75	52	56	60
3	61	44	60	45	17a	52.5	—	—	—
4	32	28	56	35	17b	63.5	—	—	—
5	33	44	32	60	18	55	48	64	55
6	33	40	20	40	19	24	16	68	35
7	46	48	68	45	20	44	36	56	30
7a	47	44	56	55	21	40	36	44	70
8	42.5	36	68	25	22	45.5	56	64	50
9	61.5	64	80	75	23	45.5	64	64	55
10	38	36	60	55	24a	54	56	60	70
11	43	—	—	—	24b	38	28	20	70
12	60	64	72	70	25a	48.5	32	32	80
13	35	48	80	55	25b	49.5	32	64	50
14	21.5	36	28	20	26	49	60	68	70
15	45.5	52	60	40					

Administration

Although attention was primarily centered upon the nine tests forming the color-form battery, all the tests described by Keehn (13) were given to 25 schizophrenic, 20 manic-depressive, and 25 neurotic males whose ages ranged from 19 to 39 years. The reason for including the extra tests was twofold. First, as the original normal group had received the whole battery it was deemed wiser to repeat this procedure with the patients rather than risk the introduction of an extra relevant variable and, second, it was anticipated that some of the extra tests would prove useful in differentiating between the groups for reasons other than their color content.

RESULTS

Table 3, by showing the percentage of *Ss* giving the color response to the individual tests, exhibits the sort of information presented by the earlier studies. Similar percentages obtained from a group of 200 normal males are also included. However, the main interest here is not centered upon the individual tests but upon the nine-item color-form test battery as a whole. Consequently,

Table 4 shows means and standard deviations for the groups where each test was given its regression weight (see Table 2) and each person's score was the algebraic sum of his weighted scores on all the nine tests. A positive score was given for a color response and a negative one was given for a form response. Again the results from the 200 normals are included.

TABLE 4

GROUP MEANS AND *SD*'S ON THE COLOR-FORM BATTERY*

Group	*N*	Mean	*SD*
Normals	200	−27.4	81.28
Schizophrenics	25	43.5	98.40
Manic-depressives	20	−2.5	90.53
Neurotics	25	−28.5	76.53

* Positive scores show a preponderance of form responses.

DISCUSSION

Normals vs. neurotics. There is widespread feeling among psychologists (10, 20), especially those having experience with the Rorschach test, that reaction to color is somehow linked with emotional instability. Consequently, since Eysenck (7), Clarke (5), and others have linked neuroticism with lack of emotional control, it seemed reasonable to expect neurotics to show greater reaction to color than normals. This was shown not to be so on the nine-test battery; neurotics reacted slightly but insignificantly less to color than the normals.

Only one of the individual tests, number 16, differentiated between the two groups. As this was at the .05 level, and as at least one such differentiation would be expected by chance, no significance was attached to this result.

Schizophrenics vs. manic-depressives. Differential reaction to color and form is one of the bases of Kretschmer's (16) typological system, and many studies carried out by his school (see Eysenck [8] for summary) suggest strongly that manic-depressives react more to color than schizophrenics. However, not only is this not supported by the scores of these groups on the present battery, but the difference, although statistically nonsignificant, is decidedly in the reverse direction.

Examination of the individual tests shows four which do differentiate between the two groups. Test 25*a*, a comparison of reaction time to color with that to shape, and test 24*b* separated the groups at the .001 level, and in both cases the manic-depressives obtained the higher color scores. Yet tests 8 and 19, differentiating between the groups at the .01 and .05 levels respectively, show the schizophrenics to react more to color. Clearly, then, the factor distinguishing between the groups is not that of reaction to color, especially as the three tests appearing in the color-form battery have the lowest weights in that battery.

One possible explanation of these anomalous findings would be that manic-depressives have a lower threshold for perceiving color in brief exposure tests (numbers 24*b* and 25*a*) but react to form when time for deliberation is allowed, as in tests 8 and 19 where the response is made with the test material still visible to *S*. However, the dangers besetting phenomenological interpretations of these tests will be apparent and no suggestion without considerable experimental backing is worth taking seriously.

Psychotics vs. normals. On the color-form test battery comparison of the normal and schizophrenic groups shows the latter to be significantly more color reactive beyond the .001 level of confidence. The manic-depressives fall between but are not significantly different from either group. Since the combined psychotic group differs significantly from the normal group also, it is impossible to say from this evidence whether the battery separates

TABLE 5

SIGNIFICANCES OF THE DIFFERENCES BETWEEN THE VARIOUS GROUPS ON THE VARIOUS TESTS

Test	Normals vs. Schizophrenics	Normals vs. Manic-depressives	Normals vs. Psychotics
4	.05		
5		.05	
7	.05		
8	.05		
9			.05
10	.05		.05
13	.001		.001
16			.05
19	.001		.001
21		.01	
24b		.003	
25a		.003	
26		.05	.05

out schizophrenics in particular, the manic-depressives behaving on this battery similarly to the normals, or whether it is measuring psychoticism in general, the two psychotic groups being combined.

In the light of Eysenck's (9) finding of a general factor of psychoticism but no group factor separating out the conventional psychotic subgroups, the latter alternative seems to be the most probable. Yet in Table 5, which depicts those individual tests differentiating between the normals and the schizophrenics, five tests occur in the color-form battery, whereas only one test in the battery distinguishes between normals and manic-depressives.

Further examination of Table 5 shows that as there is no overlap between the tests differentiating between the respective psychotic subgroups and the normals, it would appear that the postulation of a single psychotic dimension is inadequate. For if schizophrenics and manic-depressives lay along a single continuum of psychoticism, tests differentiating between the closer groups would also differentiate between the extreme groups. Whichever psychotic group lies nearer to the normals, in the present study this differentiation is not shown.

Furthermore, as the difference between the means of the schizophrenics and manic-depressives approaches significance ($t = 1.6$) even with such small numbers, it is not unlikely that large groups would differ significantly on the battery. Hence rather than conclude that reaction to color on the color-form battery is indicative of psychoticism, it would seem better to regard responses to the battery as a measure of schizophrenia.

Finally, in the last column of Table 5, six individual tests occur that discriminate significantly between the total psychotic group and the normals. Since four of these tests also differentiate between the normals and one or other of the psychotic subgroups, it is likely that in these cases the differences are largely due to the effect of the relevant subgroup. The remaining two tests, numbers 9 and 16, may be real measures of psychoticism or they may be exhibiting some of the chance differences which inevitably occur when a large number of comparisons are made, as is the case in the present investigation.

Summary and Conclusions

A battery of nine color-form tests was given to 25 schizophrenic, 20 manic-depressive, and 25 neurotic patients, and their responses were compared with each other and with those of a normal group of 200 males. Several other tests were administered at the same time. Results showed that neurotics do not differ significantly from normals in their responses to color on the test battery, and that manic-depressives respond less than schizophrenics do to color but not significantly so. The schizophrenics did, however, react significantly more to color than did the other groups, and it was argued that the battery was measuring schizophrenia specifically rather than psychoticism.

Subsidiary evidence from the individual tests showed that although schizophrenics sometimes gave significantly more color responses than manic-depressives, the reverse was true of tests involving responses to briefly exposed stimuli. It was therefore suggested that interpretation of the processes involved in reaction to these tests must not be based upon appearances. Thus although Kretschmer's school seems to be correct in its assertion that cyclothymes react more to color than schizothymes on the tachistoscopic type color-form tests, it does not follow that cyclothymes are more reactive to color in general.

Finally, from the pattern of tests differentiating between the normals and schizophrenics and the normals and manic-depressives respectively, it was argued that a single factor of psychoticism was insufficient to account for the variations between the groups. It is likely that an investigation similar to that of Eysenck (9) but including some of the present tests will result, as Eysenck suggests, in the extraction of a factor of schizothymia-cyclothymia. However, such an investigation would need to be carried out with considerable care, for if two orthogonal factors such as psychoticism and schizothymia-cyclothymia are to be posited it would be incorrect to regard psychotic subjects as necessarily at the extremes of the latter continuum.

References

1. Blomquist, I. Investigations into the color-form reaction of epileptics. *Acta Psychiat., Kbh.*, 1949, **24**, 149–166.

2. BOLLES, M. M. The basis of pertinence. *Arch. Psychol., N. Y.* 1937, No. 212.
3. BRAAT, J. P. Die experimentelle Psychologie und Kretschmer's Konstitutionstypen. *Mschr. Psychiat. Neurol.*, 1936, **94**, 273–297.
4. CAMPAILLA, G. Sulla polarizzayione percettiva verso il colore rispetto alla forma in anamalati di mente. (On perceptual polarization toward color and form in mental patients.) *Arch. Psicol. Neur. Psich.*, 1943, **4**, 1–13.
5. CLARKE, A. D. B. The measurement of emotional instability by means of objective tests. Unpublished doctor's thesis, London Univer., 1950.
6. COTTON, C. B. A study of the reactions of spastic children to certain test situations. *J. genet. Psychol.*, 1941, **58**, 27–44.
7. EYSENCK, H. J. *Dimensions of personality*. London: Kegan Paul, 1947.
8. EYSENCK, H. J. Cyclothymia and schizothymia as a dimension of personality. I. Historical review. *J. Pers.*, 1950, **19**, 123–152.
9. EYSENCK, H. J. Schizothymia-cyclothymia as a dimension of personality. II. Experimental. *J. Pers.*, 1952, **20**, 345–384.
10. FORTIER, R. H. The response to color and ego functions. *Psychol. Bull.*, 1953, **50**, 41–63.
11. GOLDSTEIN, K., & SCHEERER, M. Abstract and concrete behavior: an experimental study with special tests. *Psychol. Monogr.*, 1941, **53**, No. 2 (Whole No. 239).
12. GULLIKSEN, H. *Theory of mental tests*. New York: Wiley, 1950.
13. KEEHN, J. D. Rorschach validation III: An examination of the role of colour as a determinant in the Rorschach test. *J. ment. Sci.*, 1953, **99**, 410–438.
14. KEEHN, J. D. An experimental investigation into the validity of the use of attitude to colour and form in the assessment of personality. Unpublished doctor's thesis, London Univer., 1953.
15. KIBLER, M. Experimentalpsychologischer Beitrag zur Typenforschung. *Zbl. ges. Neurol. Psychiat.* 1925, **98**, 524–544.
16. KRETSCHMER, E. *Körperbau und Charakter*. Berlin: Springer-Verlag, 1948.
17. KUHLMANN, F. Experimental studies in mental deficiency: Three cases of imbecility (Mongolian) and six cases of feeble-mindedness. *Amer. J. Psychol.*, 1904, **15**, 391.
18. LINDBERG, B. J. *Experimental studies of color and non-color attitude in schoolchildren and adults*. Copenhagen: Levin & Munksgaard, 1938.
19. LINDBERG, B. J. Different mixtures of syntonic and asthenic personality traits. *Acta Psychiat. Kbh.*, 1950, **25**, 225–235.
20. NORMAN, R. D., & SCOTT, W. A. Color and affect: a review and semantic evaluation. *J. gen. Psychol.*, 1952, **46**, 185–223.
21. THOMSON, G. *The factorial analysis of human ability*. London: Univer. of London Press, 1950.

Received June 27, 1953.

ANXIETY AND THE MODE OF RESPONSE[1, 2]

ROBERT E. SILVERMAN[3]
Indiana University

RECENT studies in the area of perception and motivation have often produced data with contradictory and ambiguous interpretations (3, 4, 6, 8, 11, 12, 13). One explanation for this situation may be that interpretations are made which are burdened with *ad hoc* concepts like perceptual defense and perceptual sensitization that even now show signs of being eschewed by their authors (12).

It is felt that these concepts are entirely superfluous and arise because investigators fail to control the stimulus situation and to consider responses apart from any goal-seeking or drive-reducing properties. It would seem that a fruitful approach at this time would involve studying the gross responses or a general class of responses that persons make to the immediate stimulus situation. Once we have acquired some information at this level, the next step might then be concerned with investigating the development or acquisition of responses as a function of stimulus conditions.

In attempting to treat the first problem, that of studying the responses, it is necessary to develop a method of grouping or classifying responses as they relate to classes of stimuli. The classification scheme used in the experiment reported here simply entails a consideration of responses as they relate to the relevant as opposed to the incidental stimuli in a controlled stimulus situation.

Given an experimental setting, we may designate the subject's (*S*'s) task as it is structured by the experimenter (*E*) as being composed of relevant stimuli. All the controlled stimuli in the situation that have no observable relationship to the relevant stimuli are designated as incidental. We may assume that in the experimental setting an *S*'s *mode of response* can involve responding to: (*a*) the relevant or task stimuli, (*b*) the incidental stimuli, (*c*) both, (*d*) neither.

Bahrick and Fitts (2) have demonstrated that increasing the reward value of a central task interferes with performance of a peripheral task. As motivation was increased *S*s tended to respond more to the relevant as opposed to the incidental aspects of the stimulus situation. Bahrick and Fitts's *S*s were rewarded or not rewarded for each discrete response. Thus, the positive reinforcement for each response was explicitly related to the relevant task.

However, the question arises as to the general conditions that will elicit relevant as opposed to incidental responses. What, for example, will be the effect of anxiety-evoking variables? Does anxiety enhance the selectivity of responding as some authors have predicted (4, 11, 13)? Or, as a second group of writers have thought (5, 9, 10), do the conditions which evoke anxiety produce a general unselective type of responding?

The present study was designed to investigate the effects of anxiety—defined as the presentation of the threat of electric shock—on the mode of response of human *S*s. The experimental situation permitted some degree of control over the relevant and the incidental stimuli, and measures of the responses to these were obtained. If *S*'s mode of response emphasized responding to the relevant stimuli, we would expect greater improvement in his performance on the experimental task as compared to improvement in performance of those *S*s who respond to the incidental stimuli or who respond to neither the relevant nor incidental stimuli. The *S*s who respond to the incidental stimuli would manifest more incidental learning than *S*s who use the other modes of response.

The possibility exists that responding to both the incidental and relevant stimuli could raise or lower performance with respect to either class of stimuli. In order to check this, two types of relevant tasks have been used.

[1] This paper is a portion of a thesis submitted to the Department of Psychology of Indiana University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The writer is indebted to Dr. Bertram D Cohen for his advice and assistance throughout the study.

[2] A part of this paper was read before the Eastern Psychological Association at Boston in 1953.

[3] Now at New York University.

One task was considered as being difficult and the second task was a very easy one. In this way the relationship between the relevant and incidental tasks was observed.

METHOD

Subjects. The *S*s were 146 students from first-year psychology courses. They were assigned randomly to six groups. The groups each contained 25 *S*s with the exception of groups E and F, each of which consisted of 23 *S*s.

Apparatus. The experiment was performed in a sound-deadened room 8 ft. wide by 14 ft. long by 8 ft. high. The apparatus consisted of a response lever which *S* moved to indicate a response, a film strip projector which was used to present the relevant task stimuli, and a wire recorder and separate speaker attachment which were used to present the incidental stimuli. In addition, an inductorium was used to administer a sample shock to the shock-threat groups.

The response lever was a 2-in. plastic rod projecting through an aperture in a 23-in. by 43-in. wooden screen. The rod was connected to a brass slider which in turn was connected to an endless loop of string that moved a writing pen attached to an Esterline-Angus recording system. The arrangement was such that when the rod and slider were moved 9 in. the movement was recorded as a 2.2-in. move on the recording tape.

The stimuli for the relevant task were presented by a film strip projector connected to a timing system which changed the film strip frame every 10 sec. The stimuli were five black lines 0.5 in. in width and varying in height from 2.25 in. to 3.25 in. in steps of .25 in. These stimuli were presented on a 6 by 8-in. white screen at *S*'s eye level. The projector was behind the screen; thus only the screen and the response lever were visible to *S*.

The incidental stimuli were 20 two-syllable words randomly selected from 94 spondaic words developed at the Psycho-Acoustic Laboratory of Harvard University (7). The words used were: railroad, sidewalk, midway, whitewash, playmate, buckwheat, sundown, platform, lifeboat, cookbook, outside, beehive, workshop, headlight, doorstep, scarecrow, wildcat, baseball, grandson, stopwatch. Ten repetitions of the words were recorded in 5-min. intervals on a wire recorder (Radiotechnic, Model 55A) in a woman's voice. The incidental stimuli were introduced by switching on the wire recorder which played through a muffled 6-in. speaker placed in the ceiling of the experimental room. The volume output for the recorder was set at 50. At *S*'s position, the sound-level intensity of the words was approximately as loud as "subdued conversational speech."

The sample shock was administered through an inductorium energized by two 1.5-v. dry-cell batteries in series. The primary control switch was set for continuous vibrator operation and the output voltage was at maximum.

Preliminary training and pretest. Each *S* in groups A, B, C, and D was seated at the apparatus, and the projector was switched on to present the training stimulus. The *S* was instructed that this was to be a discrimination task and that he was to learn to move the key when he saw the line which was now on the screen, but that he was not to respond when any other line appeared. A brief training series followed during which *E* reinforced a correct response by saying "that's right" and indicated a wrong response (failing to move when the correct line was presented or moving the response key when an incorrect line was presented) by stating "you should not move the key since that is a longer (shorter) line" or "you should move the pointer since that is the correct line." During the training series the stimuli were presented in the following order: 3, 3, 3, 3, 5, 3, 1, 3, 2, 3, 4, 3, where 3 indicates the correct line (2.75 in.), 1 is the longest line (3.25 in.), and 5 is the shortest line (2.25 in.). Following the training series each *S* received a pretest in which the lines were presented in the following order: 3, 5, 3, 3, 3, 2, 3, 4, 3, 1, 3, 2, 3, 3, 3, 4, 3, 1, 3, 5.

An *S* who made more than five errors on the pretest was eliminated from the remainder of the experiment. Eleven *S*s were eliminated. An error was defined as responding incorrectly to a line other than 3 or failing to respond to line 3.

The *S*s in groups E and F were exposed to a very easy discrimination task during the training and pretest series. They were instructed to respond whenever a line (any line) appeared on the screen. The lines were the same as those used for the other groups.

Test series. For all *S*s the pretest series was followed by a 1-min. rest period and then a test series. In the test series the order of presentation was: 3, 5, 1, 2, 4, 5, 4, 1, 2, 3, 4, 1, 2, 5, 3, 1, 4, 3, 1, 5, 2, 3, 4, 4, 5, 1, 3, 2, 5, 4, 1, 4, 2, 1, 5, 3. The *S*s in groups E and F again merely had to respond when any line appeared on the screen.

The experimental conditions were introduced during the test series for all six groups. Group A performed the test series while being threatened with electric shock. Electrodes were attached to *S*'s left wrist and left ankle (the ankle electrodes were "props" in that they were not attached to the inductorium). Each *S* was told that he would receive an occasional electric shock during the test series. A sample shock was given before the series began, but shock was not administered again during the experiment. One minute after the presentation of the first line, the words (incidental stimuli) were started. The words continued until the final 15 sec. of the test series. Thus the words were presented while visual stimuli 7 through 34 were being presented.

TABLE 1

MEANS AND STANDARD DEVIATIONS OF THE RECALL AND RECOGNITION SCORES

Measure	Difficult		Easy	
	A Shock	B Non-shock	E Shock	F Non-shock
Recall				
Mean	6.56	12.20	11.09	19.13
SD	6.93	11.83	10.44	11.61
Recognition				
Mean	4.31	6.19	5.09	7.25
SD	4.60	4.73	4.18	3.88

TABLE 2

ANALYSIS OF VARIANCE OF THE RECALL AND RECOGNITION SCORES

Source	Recall				Recognition			
	Sum of Squares	*df*	Mean Square	*F*	Sum of Squares	*df*	Mean Square	*F*
Difficulty of task	786.75	1	786.75	5.879*	20.25	1	20.25	1.010
Shock and nonshock	1105.71	1	1105.71	8.263**	97.35	1	97.35	4.856*
Interaction task × Shock	34.58	1	34.58	.258		1	.47	.023
Within cells	12311.34	92	133.82		1844.25	92	20.05	

* Significant at the .05 level of confidence.
** Significant at the .01 level of confidence.

Group B performed the discrimination test under the same conditions as group A with the exception of the shock threat. Group C performed the discrimination test under the threat of shock as did group A, but this group was not exposed to the words. Group D was exposed to neither the shock threat nor the words. Groups E and F performed the easy discrimination. Group E was subjected to shock threat and exposed to the words during the test series. Group F received no threat of shock but was exposed to the words during the test series.

Test for incidental learning. Following the discrimination test, *Ss* who had been exposed to the words were given a recall test and a recognition test for the incidental words. Recall scores were obtained by allowing full credit for a word correctly recalled, half credit for one syllable of the two-syllable word correctly recalled, and one-fourth credit for any two-syllable, spondaic word designated. The total score was multiplied by 4 to eliminate fractions. Recognition scores were obtained by having *Ss* select words from a matrix containing 80 words. A score was corrected for guessing by subtracting from the number correct one-third the number of incorrect words.

RESULTS

Incidental learning. The mean recall and recognition scores for the groups exposed to the words are presented in Table 1. Inspection of these tabulated data indicates more evidence of incidental learning in the groups (E, F) which performed the easy discrimination task than in the groups (A, B) which performed the more difficult discrimination task. Even more marked is the superiority in incidental learning, as evidenced by their mean scores, of the nonshock groups (B, F) over the shock groups (A, E).

TABLE 3

MEANS AND STANDARD DEVIATIONS OF THE DISCRIMINATION TEST SCORES

Measure	Words		No Words	
	A Shock	B Nonshock	C Shock	D Nonshock
Mean	27.96	29.56	30.00	27.00
SD	4.29	3.24	3.44	3.21

TABLE 4

ANALYSIS OF VARIANCE OF THE DISCRIMINATION TEST SCORES

Source	Sum of Squares	*df*	Mean Square	*F*
Words	1.69	1	1.690	.127
Shock	12.25	1	12.250	.928
Words × Shock	132.25	1	132.250	9.941**
Within cells	1277.12	96	13.303	

** Significant at the .01 level of confidence.

The results of the analysis of variance for the incidental learning scores are presented in Table 2.[4] That the degree of difficulty of the discrimination task was an effective variable is indicated by the *F* ratio for the recall scores. However, the differences between the nonshock and shock conditions were statistically significant for both the recall and recognition scores.

Discrimination task. Discrimination scores were obtained by tabulating the number of correct responses made by each *S* during the pretest and test series. An analysis of the pretest data indicated that significant differences did not exist among the four groups. Thus it

[4] The variances for the recall scores were not homogeneous. Transforming the data so as to satisfy the assumption of homogeneity does not change the probabilities (*F* for shock, nonshock was 8.779; *F* for difficult, easy was 5.035). Therefore the statistical test for the nontransformed data is presented in Table 2.

was assumed that the groups were relatively similar with respect to initial discrimination ability.

The means and *SD*'s for the discrimination-test scores are presented in Table 3. The analysis of variance for these data is presented in Table 4.

A significant interaction effect was found between the experimental conditions where no over-all differences had been found between the words, no-words conditions or between the shock, nonshock conditions. This interaction would seem to indicate that the two experimental conditions, shock and words, act differently in combination than either of these conditions acts alone. The discrimination of *S*s exposed to shock threat is slightly impaired when the words are present as compared to the no-words condition. However, this impairment cannot be attributed merely to the presence of the words, for a comparison of the words and no-words groups which are not exposed to shock threat reveals that the words condition alone improved performance. It is only when the words are combined with the threat of shock that performance is disrupted.

DISCUSSION

The results of the incidental learning tests are generally in agreement with those reported by Aborn (1) who used very different stimulus materials. His data indicate that "ego threat" decreased incidental learning as measured by recall and recognition scores. We might interpret such findings as offering support for the notion that threat acts to increase the relevant mode of response. In the study reported here, we would expect an *S* who responds predominantly to the relevant stimuli to show decreased incidental learning, as would an *S* who responds to neither class of stimuli. Responding to both classes should produce relatively good performance in both spheres since the facilitating effects of the words would cancel the interfering effects of the discrimination task. It is also possible that responding to some degree in the direction of the incidental stimuli might increase the relative efficiency in both tasks.

The mode of response of *S*s exposed to shock threat may be considered as being directed toward the relevant stimuli since these *S*s show relatively good discrimination when no words are present and poor discrimination as well as poor incidental learning when exposed to the words. The possibility exists that the shock-threat *S*s attempt to avoid listening to the words and that these avoidance responses in turn interfere with discrimination accuracy. This bit of speculation stems from the fact that the group (A) exposed to both words and shock threat shows impaired discrimination.

The fact that shock threat elicits a relevant mode of response is additional evidence for what appears to be the selective and stereotyped patterns of response often attributed to the chronically anxious person. Whether or not the condition of anxiety per se enhances this mode of response is problematical. However, by relating the mode of response to the discrete responses investigated by Bahrick and Fitts (2) the acquisition of the relevant response can be investigated within a general learning framework. We may consider the mode of response as a learned response acquired in the same manner as any class of discrete responses. Accordingly, we would predict that persons with a history of clinical anxiety would in a situation similar to that reported above manifest a strong relevant mode of response.[5]

In the study reported here, the relevant response is defined by *E*. The task itself cannot be said to have any anxiety-reducing properties, but by responding to the structured task *S*s are literally conforming in a miniature social situation. This is to be expected, since, in a discussion of human behavior, relevance in its broad sense is defined by society. Thus we would predict and do find that a large segment of the anxious patients are rigid conformists. The typical compulsive patient with symptoms of manifest anxiety is largely a relevant responder although the rituals and compulsions often indulged in by these patients appear to be unusual. The explanation for this is that relevance is defined by subcultures and even by fantasy activity as well as by the mores of society at large. Thus a patient who indulges in a bizarre ritual is probably making his own interpretation of relevance.

[5] A study using test scores to establish a history of anxiety is in progress to test this prediction.

SUMMARY

This experiment investigated the influence of shock threat on the responses to incidental auditory stimuli and relevant visual stimuli. The *S*s were trained to perform a visual discrimination task and were then tested for discrimination accuracy. The *S*s who performed the discrimination test in the presence of incidental auditory stimuli were tested for recall and recognition of these stimuli.

The results indicated that *S*s exposed to the threat of electric shock manifested less incidental learning than *S*s not exposed to shock threat. The shock-threat condition in the presence of the incidental stimuli impaired discrimination, but the shock-threat condition alone improved the accuracy of discrimination.

The results were interpreted as lending support to the clinical data that show anxious persons to be relevant responders or conformists.

REFERENCES

1. ABORN, M. The influence of experimentally induced failure on the retention of material acquired through set and incidental learning. *J. exp. Psychol.*, 1953, **45**, 225–231.
2. BAHRICK, H. P., FITTS, P. M., & RANKIN, R. E. Effect of incentives upon reactions to peripheral stimuli. *J. exp. Psychol.*, 1952, **44**, 400–406.
3. BRUNER, J. S., & GOODMAN, CECILE C. Value and need as organizing factors in perception. *J. abnorm. soc. Psychol.*, 1947, **42**, 33–44.
4. BRUNER, J. S., & POSTMAN, L. Emotional selectivity in perception and reaction. *J. Pers.*, 1947, **16**, 69–77.
5. CAMERON, N. *The psychology of the behavior disorders.* New York: Houghton Mifflin, 1947.
6. CARTER, L. F., & SCHOOLER, K. Value, need, and other factors in perception. *Psychol. Rev.*, 1949, **56**, 200–207.
7. DAVIS, H. (Ed.) *Hearing and deafness, a guide for laymen.* New York: Murray Hill Books Inc., 1947.
8. HOWES, D. H., & SOLOMON, R. L. Visual duration thresholds as a function of word probability. *J. exp. Psychol.*, 1951, **41**, 401–410.
9. JENKINS, J. G. Instruction as a factor in "incidental" learning. *Amer. J. Psychol.*, 1933, **45**, 471–477.
10. LIDDELL, H. The role of vigilance in the development of animal neurosis. In P. Hoch & J. Zubin (Eds.), *Anxiety.* New York: Grune & Stratton, 1951. Pp. 183–196.
11. MCGINNIES, E. Emotionality and perceptual defense. *Psychol. Rev.*, 1949, **56**, 244–251.
12. POSTMAN, L., BRONSON, WANDA C., & GROPPER, G. L. Is there a mechanism of perceptual defense? *J. abnorm. soc. Psychol.*, 1953, **48**, 215–224.
13. VANDERPLAS, J. M., & BLAKE, R. R. Selective sensitization in auditory perception. *J. Pers.*, 1949, **18**, 252–266.

Received July 2, 1953.

AN ANALYSIS OF THE CATTELL-LUBORSKY HUMOR TEST INTO HOMOGENEOUS SCALES

JAMES K. YARNOLD AND MARVIN H. BERKELEY

Human Resources Research Center, Lackland Air Force Base

THE ANALYSIS of reactions to various types of humor stimuli has sometimes been regarded as a potentially valuable means of personality measurement. The problem of classification and interpretation of responses to humor, however, has been a deterrent to the exploratory use of humor scales in personality research.

Andrews accomplished the first large-scale statistical analysis of a sample of humor items (2). His factor analysis of a 24-item matrix demonstrated that humor response was not of an entirely general nature as maintained by certain early theorists (9). The item intercorrelation range (−.10 to .59, tetrachoric) indicated the magnitude of covariation. In addition, his study showed that the specific item structure, whether it be cartoon, joke, or limerick, was not of primary importance. While this latter information is certainly important, it should be pointed out that Andrews' factors would have little utility as actual measures of humor response.

Similarly, the research of Cattell and others has yielded little in terms of sufficiently homogeneous and independent measurement devices. The first clusters (Cattell and Luborsky [4]) were developed with a too lenient criterion for including an item in a cluster, resulting in measures with low reliabilities. A later study by Cattell, Luborsky, and Hicks (6) was concerned with the factor analysis of a sample of humor items. The factor tests resulting from this analysis also fail to meet the minimum qualifications for an adequate set of scales. Moreover, their psychological meaning appears overly complex. The median reliability (Kuder-Richardson Formula 20) is .29, the coefficients ranging from .18 to .51. Although the median intercorrelation between factor tests is only .17, the limiting effect of the very low reliabilities suggests that this degree of independence may be spurious.

On the basis of the scaling deficiencies of the humor clusters and factors derived from the three major attempts at empirical classification of humor items, it is apparent that no adequate measures of humor response yet exist. The task of the project described here has been to derive an unspecified number of homogeneous scales from a heterogeneous pool of humor items.

RESEARCH DESIGN

The research design for this task was organized as follows: (*a*) selection of first item set, (*b*) selection of subject sample, (*c*) test administration, (*d*) analysis into tentative scales, (*e*) selection of second item set, (*f*) analysis into final scales, (*g*) comparison of the tentative and final tests, and (*h*) psychological interpretations of final scales. Each of these phases of the research design is discussed in the above order.

Selection of First Item Set

It is obvious that the adequacy of coverage of the personality domain is limited by the coverage of the items selected for analysis. Cattell and Luborsky had selected the items for their C-L Humor Test, Forms A and B, from an extensive pool of 2,000 jokes and limericks. It was decided to capitalize on their previous work by selecting as the initial set one of the two forms of their test. Form A (91 items) required the forced-choice selection of the funnier of two jokes. Form B (112 items) required a response of *funny* or *dull* for each item. The first approximation of homogeneous scales was based on the collection of the more highly intercorrelated clusters developed by Cattell and Luborsky (5). Since Form A clusters (forced-choice items) presented many more interpretative problems than Form B clusters, items of the latter form were selected as the initial item pool.

Selection of Subject Sample

The sample on which the Form B homogeneous keys were developed consisted of 500 completed answer sheets of officer candidates at Lackland Air Force Base. The test was administered while most of the officer candidates were in their second week of training. In order to facilitate computational procedures, the sample was decreased to 500 cases from a number slightly over that amount, and stratified on the basis of sex.

The cross-validation sample used to test the stability of the Form B keys consisted of 293 officer candidates of a subsequent class. The proportion of males and females as well as other OCS selection procedures remained unchanged for the cross-validation sample.

TABLE 1

PRODUCT-MOMENT INTERCORRELATIONS AND DISTRIBUTIONAL STATISTICS OF THE FINAL HOMOGENEOUS TESTS WITH AN INDEPENDENT SAMPLE

(Sample: 230 Students, OCS Class 53-A)

Test	RF_2	RG_1	RG_2	R	S	T	Mean	*SD*	K.R. 20*	No. of Items
RD_1	.00	−.09	.31	.06	.19	.00	13.35	3.46	.70	20
RF_2		.27	.11	.45	.23	.29	6.22	2.37	.43	15
RG_1			.26	.42	.33	.47	4.27	2.65	.62	15
RG_2				.15	.33	.24	6.49	2.02	.45	11
R					.27	.43	4.91	1.88	.42	10
S						.32	4.44	1.83	.40	9
T							3.40	2.06	.53	11

* Kuder-Richardson Formula 20.

Test Administration

The Form B item set was administered in testing rooms containing approximately 40 subjects (*Ss*) each. The *Ss* were told that their "sense of humor" was being measured. (Cattell and Luborsky had labeled Form B as "A Humor Rating Quiz.") An unlimited amount of time was allowed for test completion, although *Ss* were asked to record their immediate opinion of items. In order to reduce social facilitation, *Ss* were requested to inhibit audible laughter. However, in all testing groups, a few *Ss* found it impossible to comply with this request. Since *Ss* answered items at different rates of speed, it seems unlikely that individual item responses were seriously affected.

Analysis into Tentative Scales

A direct and economical means of developing independent tests that have an adequate degree of discriminating power has been suggested by DuBois, Loevinger, and Gleser (8). This technique is known as maximizing the discriminating power of a test. This method avoids a number of difficulties inherent in the application of factor analytic techniques to items and also appears to lack the principal limitations of other methods of maximizing test homogeneity (8, pp. 2–6). This technique was thus selected for clustering the set of humor items into homogeneous tests. The major steps are listed below:[1]

Grouping of items into initial submatrices. It seemed that the later clusters developed by Cattell and Luborsky (5) might possibly be better than purely a priori groupings of items. Thus, Cattell-Luborsky cluster scores for OCS Class 51-A were intercorrelated, and on the basis of the intercorrelations, the C-L clusters were grouped into three matrices containing about 40 items each.

Analysis into tentative scales. Five tentative scales were developed by the DuBois, Loevinger, and Gleser method for maximizing the discriminating power of a test. These scales and others, subsequently derived, will henceforth be referred to by the letter of the matrix from which they were evolved and a subscript corresponding to the order of evolvement. The tentative scales were labeled D_1, F_2, F_3, G_1, and G_2.

Selection of second item set. The second item set from which the tentative scales were to be expanded into final scales consisted of the 91 forced-choice items of Form A of the C-L Humor Test. However, *Ss* were not required to choose one of two paired jokes, but instead made an independent rating of *funny* or *dull* for each joke. This resulted in a pool of 182 items which, along with Form B, was administered to a new sample of 400 officer candidates.

Analysis into final scales. It was the purpose of this phase of the research to increase the discriminating power of the tentative scales and to develop additional scales from items of a new sample. In this analysis four of the five tentative scales were retained and increased in length. These scales were relabeled RD_1, RF_2, RG_1, RG_2, the R standing for revised. Three additional scales were evolved with labels R, S, and T. The seven scales constituted the final homogeneous tests. Their intercorrelations and distributional statistics are given in Table 1, for an independent sample of 230 officer candidates.

Psychological Interpretations of Final Scales

On the basis of a number of experiments Sears (14) concluded that a joke may be analyzed into two central elements: thematic and schematic. The schematic element is the technical structure or form of the joke, while the thema "constitutes the meaningful human content, or plot, of a joke." It is the joke thema "which evokes the asocial or repressed needs, principally Aggression, Superiority, and Sex. Such a joke is the imaginal presentation of a thema which provides—suddenly and gratuitously—the end situation of one of the subject's needs, fused with n Play. The mirth-thema, then, brings added delight by satisfying in

[1] The details of the procedures and the results have been filed with the American Documentation Institute. Order document No. 4286 from the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.75 for 35 mm. microfilm or $2.50 for photoprints. Make checks payable to Chief, Photoduplication Service, Library of Congress.

fantasy some purely egocentric and usually hostile wish. The delight is not inhibited because it can escape, disguised by the playful social feelings engendered by the mirth-schema in which the mirth-thema is embedded" (13, p. 584).

A tentative description within the Sears frame of reference of each of the final homogeneous scales is presented below. A typical item is presented with each test for illustrative purposes.

RD_1. The element common to the items of RD_1 is that of sexual play or sexual need. Sexual fulfillment is accepted as an integral part of life. However, this need for sexual fulfillment results in frustrations which are not necessarily resolved by "conventional" social mores.

It seems reasonable to hypothesize that RD_1 will be highly weighted on the following factors or scales: "Social Introversion" (Guilford & Guilford [11]), "Adventurous Cyclothymia vs. Inherent Withdrawn Schizothymia" (Cattell [3]), and "Social Introversion" (Mosier, [12]).

Example: The census taker viewed Mrs. Jones and the six tots of varying ages around her with a puzzled frown. He seemed particularly interested in a squirming infant in her arms. "I don't quite understand you," he admitted. "Did you say that your husband died six years ago?" "Yes, sir," she replied emphatically. "*He* died, but I didn't."

RF_2. The items of this test have in common the display of active self-assertion by an independent, dominant, and somewhat unconventional individual.

It is hypothesized that RF_2 affords a subtle measure of the "Dominance" factor reported by Cattell (3), the "Lack of Agreeableness" factor of Guilford and Guilford (11), or inversely, the "Sociable" factor of Thurstone (15).

Example: One prehistoric man to another: "Now that we've learned to communicate with each other—shut up!

RG_1. In all items of this test, one detects the presence of a scheming, manipulative individual who is directing the action of his associates and taking full advantage of all opportunities to assert himself. This assertion usually takes aggressive form of the extrapunitive type. In some cases the aggression is explicit, as in a hard, cutting retort to a rather stupid, dull, or otherwise uncalled for remark by an overly inquisitive, annoying, or generally dull character. In other items the aggression is implicit, as in the case of the clever, rather malicious opening statement designed to allow the replying individual to make himself appear ridiculous. In all cases the directing influence of a shrewd, exploitative, and somewhat aggressive individual is evident.

It is suggested that RG_1 may be related to the variance which the 16PF Questionnaire Factor L is purported to measure, "Paranoid Schizothymia vs. Trustful Accessibility" (5).

Example: A famous musician once asked a man why he forever sang one tune.

"Because it haunts me," replied the man.

"No wonder," said the musician, "you are continually murdering it."

RG_2. The items of RG_2 are characterized by a common element consisting of a general lack of social insight and sophistication on the part of one character. In nearly all cases a naive individual (usually juvenile) makes a statement which has the effect of upsetting the equilibrium or poise of the other individual.

Example: Willie: "Papa, why don't you go away so I can see Mama dump that salesman off the fire escape?"

Papa: "Why, Willie, what on earth are you talking about?"

Willie: "Well, I heard her tell him that she'd tip him off as soon as you left."

R. The central thema of this test deals with the ridicule of authority figures (i.e., judges, ministers, doctors, teachers, parents, etc.).

It is hypothesized that individuals obtaining a low score on Test R are those who identify with the ridiculed authority figure and are prototypical of the authoritarian personality (Adorno *et al* [1]).

Example: "I rise for information," said a member of the legislative body.

"I'm very glad to hear it," said a bystander, "for no man needs it more."

S. The central element of Test S appears to be a passive acceptance of inferior or socially disapproved behavior by the leading character. In each case the individual reveals his own intellectual or moral inertia. There is no attempt to alter behavior in a more desir-

able direction. Actually, the principal character often seems satisfied with his resignation. The general implication is that the individual is too weak to better himself and so merely accepts his intellectual or moral inferiority.

Example: An intoxicated man, walking along the street, regarded the moon with much contempt. "You needn't feel so proud," he said, "you are full only once a month, and I am full every night."

T. The humor of this test seems to have its basis in the slightly cynical realism displayed by the principal characters. It is hypothesized that high scorers on Test T will be those individuals who are sensitized to the foibles of men or women in everyday real-life situations. The common element of Test T is one which deals with the contrast between human pretension and performance. It appears that this test will relate to the emotional stability vs. general neuroticism factor suggested by Eysenck (10) and Cattell (3), and to the lack of objectivity factor as suggested by Guilford (11).

Example: "Did he take his misfortunes like a man?"
"Precisely. He laid the blame on his wife."

CONCLUSIONS

The following conclusions are drawn from this study:

1. An analysis of both Form A and Form B of the Cattell-Luborsky Humor Test by a recently proposed cluster method has resulted in the development of seven scales which appear to be superior, at least statistically, to the scales previously derived for this test.

2. The apparent content of each of the seven scales has been briefly described and related to other factors and scales.

3. The analysis of reactions to humor stimuli does seem to provide a potentially effective and subtle means for the measurement of personality; however, validation is necessary.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality.* New York: Harper, 1950.
2. ANDREWS, T. G. A factorial analysis of responses to the comic as a study in personality. *J. gen. Psychol.*, 1943, **28**, 209–224.
3. CATTELL, R. B. *Personality, a systematic theoretical and factual study.* New York: McGraw-Hill, 1950.
4. CATTELL, R. B., & LUBORSKY, L. B. Personality factors in response to humor. *J. abnorm. soc. Psychol.*, 1947, **42**, 402–421.
5. CATTELL, R. B., & LUBORSKY, L. B. *Handbook for the C-L Humor Test of Personality.* Champaign, Ill.: Institute for Personality and Ability Testing, 1947.
6. CATTELL, R. B., LUBORSKY, L. B., & HICKS, V. *Revised handbook for the C-L Humor Test of Personality.* Champaign, Ill.: Institute for Personality and Ability Testing, 1950.
7. CATTELL, R. B., SAUNDERS, D. R., & STICE, G. *Handbook for the 16 PF Test.* Champaign, Ill.: Institute for Personality and Ability Testing, 1951.
8. DUBOIS, P. H., LOEVINGER, JANE, & GLESER, GOLDINE C. The construction of homogeneous keys for a biographical inventory. *USAF, Hum. Resour. Res. Cent., Res. Bull.*, 1952, No. 52-18.
9. EASTMAN, M. *The sense of humor.* New York: Scribner's, 1921.
10. EYSENCK, H. J. *Dimensions of personality.* London: Kegan Paul, 1947.
11. GUILFORD, J. P., & GUILFORD, RUTH B. Personality factors *S*, *E*, and *M*, and their measurement. *J. Psychol.*, 1936, **2**, 109–127.
12. MOSIER, C. I. A factor analysis of certain neurotic tendencies. *Psychometrika*, 1937, **2**, 263–286.
13. MURRAY, H. A. *Explorations in personality.* New York: Oxford Univer. Press, 1938.
14. SEARS, R. N. Dynamic factors in the psychology of humor. Unpublished doctor's dissertation, Harvard Univer., 1934.
15. THURSTONE, L. L. The dimensions of temperament. *Psychometrika*, 1951, **16**, 11–20.

Received July 6, 1953.

SOME EFFECTS OF UNEQUAL DISTRIBUTION OF INFORMATION UPON GROUP PERFORMANCE IN VARIOUS COMMUNICATION NETS[1]

MARVIN E. SHAW[2]

University of Wisconsin[3]

SINCE Bavelas (1) first presented a mathematical model for group structures (or communication nets), several studies (2, 4, 5) have been concerned with finding behavioral correlates of the various measures which the model generates. The measure that has so far proved most significant is *centrality*. An index of centrality for any particular group structure is obtained by first counting the minimum number of links (communication channels) which must be crossed in order for a given position in the net to communicate with every other position in the net. This count is called the *sum of distances* for that position. *Individual centrality* is then computed by dividing the total sum of distances for the entire net by the sum of distances for the individual position, and *net centrality* is computed by summing the individual centralities in the net.

Centrality, according to Leavitt (5), correlates well with the behavior of people in different kinds of structures. He required groups to solve simple problems and compared their performance in different nets. Centrality accounted reasonably well for differences in accuracy, total activity, satisfaction of group members, emergence of a leader, and organization of the group. In Leavitt's study, however, the information from which the group solved its problems was distributed equally among members of the group. It is a reasonable hypothesis that distribution of information, as well as centrality, may be an important variable in group behavior. The purpose of the present study is to examine such a hypothesis.

The reasoning was as follows: centrality correlates with problem-solving behavior because the greater the centrality of a position the more information is available to the individual in that position. If this be so, then distributing information unequally in the first place by giving one position more information than another ought to have the same consequences as increasing the centrality of that position.[4] Therefore, an experiment was designed in which centrality could be studied simultaneously with information distribution in order to evaluate the relative contributions and interactions of the two variables.

METHOD

Apparatus and materials. In the present study three four-cell nets were selected for investigation: the wheel, the slash, and the circle. These nets are represented diagrammatically in Fig. 1, and their chief characteristics are given in Table 1. These nets were selected because the net centralities progress in equal steps from 16.0 for the circle to 16.8 for the wheel.

The apparatus consisted of a room which was partitioned into four cubicles. Each cubicle was connected to every other cubicle by means of a slot through which message cards could be passed. These slots were fitted with overlapping cloth "curtains" so that it was impossible for one subject (*S*) to see into another's cubicle. The frames supporting these curtains were painted to indicate the cubicle into which the slot opened. Each cubicle was supplied with an electrical switch which, when thrown, turned out a light on *E*'s signal panel.

The following is an example of the type of problem used:

> Statement of problem:
>
> A small company is moving from one office building to another. It must move *four* kinds of equipment: (1) Chairs, (2) Desks, (3) Filing Cabinets, and (4) Typewriters. How many trucks are needed to make the move in one trip?

[1] This research was supported in whole or in part by the United States Air Force under contract AF 18(600)–340, monitored by the Human Resources Research Institute, Maxwell Air Force Base, Alabama.

[2] The present paper represents a portion of the dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of Wisconsin, 1953. The writer wishes to express his appreciation to Dr. J. C. Gilchrist, under whose supervision this investigation was conducted, for his guidance, helpful suggestions and constructive criticisms in every phase of the investigation.

[3] Now at The Johns Hopkins University.

[4] The assumption that increased information is equivalent to increased centrality is undoubtedly an oversimplification of a rather complex process. Altering the proportions of the total information assigned to the various positions almost certainly changes other characteristics of the net as well as centrality. However, this oversimplification is perhaps justified in this early stage of theoretical development.

Items of information needed to solve:

1. The company owns a total of 12 desks.
2. The company owns a total of 48 chairs.
3. The company owns a total of 12 typewriters.
4. The company owns a total of 15 filing cabinets.
5. One truck will carry 12 typewriters and nothing else.
6. One truck will carry 3 desks and nothing else.
7. One truck will carry 5 filing cabinets and nothing else.
8. One truck will carry 24 chairs and nothing else.

Two other problems of the same type were employed. One of these required the *S*s to determine the time of arrival of a plane at a specified town, and the other required them to determine the amount of gasoline required to move vehicles from one motor pool to another. For each of these three problems, the statement of the problem was typed completely on each of four separate cards and each item of information was typed on a separate card. The materials for a single problem consisted of these twelve white cards.

Subjects. The *S*s were students enrolled in psychology or sociology courses at the University of Wisconsin. Ninety groups (sixty exclusively female and thirty exclusively male) were run. Each of these groups was assigned at random to one of six experimental conditions, subject to the restriction that five groups of males and ten groups of females be assigned to each experimental condition. The six experimental conditions represented the possible combinations of two types of information distribution and three patterns of communication.

Procedure. Each group was required to solve the three problems mentioned above. Under conditions of equal distribution of information, for each problem each *S* was given one of the problem cards and two of the information cards. Under conditions of unequal distribution of information, the *S* in Position A was given one of the problem cards and five of the information cards, and each of the other *S*s was given one of the problem cards and one of the information cards. The distribution of particular items among positions was randomized by shuffling the white cards before each session. The sequence in which the problems were presented was varied according to a systematic latin square such that each of the three problems was given first, second, and third an equal number of times, and each problem preceded and followed every other problem an equal number of times.

With the exception of communication net and distribution of information, the procedure was exactly the same for all groups. Each of the four *S*s was brought into the experimental room separately, seated according to a randomized schedule, and requested to remain quiet. The *S*s were told the "names" of the members of the group (i.e., the color used to identify each member), the ones with whom each could communicate, the method of communication, and the general nature of the tasks.

When all three problems had been completed, each *S* was asked to fill out a three-part questionnaire. Part 1 required sociometric choices of the member of the group most preferred for six different activities ranging from one almost identical to that just completed to one almost completely unrelated to the activity just completed. Part 2 required each *S* to indicate on a nine-point scale how well he liked his job in the group, how he would compare his position with each of the other positions, how well he thought the group members cooperated with each other, and how he would rate the performance of the group. Part 3 required an answer to the question, "Did your group develop a system ——? If so, describe it briefly."

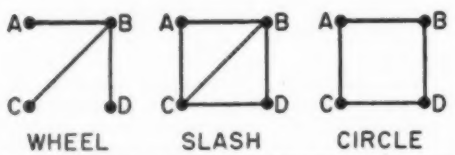


FIG. 1. DIAGRAM OF THE THREE COMMUNICATION NETS TO BE INVESTIGATED

TABLE 1

CHARACTERISTICS OF THE EXPERIMENTAL NETS

Net	Number of Channels	Sum of Distances	Net Centrality	Individual Centralities				Maximum Differences in Individual Centralities
				A	B	C	D	
Wheel	3	18	16.8	3.6	6.0	3.6	3.6	2.4
Slash	5	14	16.4	3.5	4.7	4.7	3.5	1.2
Circle	4	16	16.0	4.0	4.0	4.0	4.0	0.0

RESULTS

The results of this study fall under four general headings: (*a*) time scores, (*b*) communication units, (*c*) errors, and (*d*) questionnaire results. The first three were taken as indicators of efficiency; the questionnaire responses were used as indicators of morale and leadership.

Time scores. Time to reach a solution can be measured in at least three ways: (*a*) the fastest single correct trial, (*b*) the average time per problem per position, and (*c*) the total elapsed time per problem. Analyses of variance applied to the results using the first two time measures revealed no significant differences. The means of total elapsed time per problem are presented in Table 2. Analysis of variance showed significant differences between trials ($p < .001$). This probably represents learning to use the net, since all the problems used were well within the experiences of the *S*s. There were no over-all differences between nets or between distribu-

TABLE 2
MEANS OF TOTAL ELAPSED TIME PER PROBLEM IN MINUTES

Trials	Equal Distribution			Unequal Distribution		
	Wheel	Slash	Circle	Wheel	Slash	Circle
1	16.12†	16.22†	17.71†	16.29†	16.57†	16.05†
2	11.68	10.98	9.30	11.57	11.31	10.10
3	10.09	8.52*	8.26*	9.30	8.61*	7.77*

* Separated as stragglers.
† Separated as a homogeneous group.

TABLE 3
MEAN TIME REQUIRED BY EACH POSITION IN MINUTES

Net	Positions			
	A	B	C	D
Wheel				
Equal	11.06	9.73*	11.23	11.14
Unequal	8.61‡	10.91	10.54	11.71
Slash				
Equal	10.50	10.02	9.94	10.43
Unequal	8.56‡	10.17	10.73	11.55*
Circle				
Equal	10.37	10.21	9.98	9.93
Unequal	7.85‡	10.13	9.75	10.05

* Separated as stragglers.
‡ Separated by a significant gap.

TABLE 4
MEAN NUMBER OF ITEMS REQUIRED TO REACH A SOLUTION BY EACH NET

Trials	Equal Distribution			Unequal Distribution		
	Wheel	Slash	Circle	Wheel	Slash	Circle
1	53.5*	74.9	76.7	51.4*	82.6*	72.2
2	58.0	70.1	59.9	48.4*	76.9	68.8
3	44.9*	62.5	60.8	43.3*	65.3	58.1

* Separated as stragglers.

tions of information. However, the nets × trials × distribution interaction was significant ($p < .01$).

In order to determine as nearly as possible just where the differences occurred, Tukey's gap test (7) was applied to the mean scores.[5] The results of this test are indicated in Table 2 by the symbols placed beside the separated means. The scores for the first trial differed as a group from those for all other trials, indicating that most of the learning occurred during the first trial. On the last trial, the slash and the circle required less time to reach a solution than did the wheel, regardless of information distribution. However, a precise relationship between net centrality and time was not found.

The mean time required by each position is shown in Table 3. Analyses of variance for each net taken separately yielded essentially the same results for all three nets. Distribution as a main effect was not significant, but positions and the groups × positions interaction were both significant ($p < .001$) for all three nets. The results of the gap test are indicated by symbols. Under conditions of equal distribution of information, only Position B in the wheel was separated. This means that only relatively large differences in individual centrality produce differences in time required to reach a solution. Under conditions of unequal distribution of information, the position receiving the most information (Position A) required significantly less time to reach a solution than did any other position.

Communication units. Communication units were measured by counting the number of separate items transmitted by the *S*s in a given net or position, since it was possible for one card to contain several such items.[6] An item is here defined as any simple sentence, or any meaningful part of a complex or compound sentence. Table 4 presents the mean number of items transmitted by each net. Analysis of variance indicated that there were significant net differences ($p < .001$), trial differences ($p < .001$), and second-order interaction effects ($p < .001$). Examination of Table 4 reveals that the chief difference between nets was that the wheel required fewer items to reach a solution than did the slash and the circle. Generally, the number of items per trial decreased with practice in much the same way as did the time scores. The significant interaction is apparently due to the relatively large number of items transmitted by the wheel (equal distribution) on Trial 2, and by the slash (unequal distribution) on Trial 1. For items transmitted, as

[5] In this test and in all subsequent applications of the gap test, the .05 level of confidence was accepted.

[6] Results of an analysis of number of cards used were essentially the same as the results reported here.

TABLE 5

MEAN NUMBER OF ITEMS COMMUNICATED PER PROBLEM FOR EACH POSITION

Net	Positions			
	A	B	C	D
Wheel				
Equal	7.8	27.4‡	9.2	7.9
Unequal	10.9*	24.0‡	6.3	6.6
Slash				
Equal	14.9†	19.6	20.6	14.1†
Unequal	19.5	21.2	20.4	13.7†
Circle				
Equal	15.8	16.7	17.4	15.6
Unequal	18.8	15.4	16.8	15.7

* Separated as stragglers.
† Separated by a significant gap as a homogeneous group.
‡ Separated by a significant gap.

for the time scores, net centrality was found to be an inadequate index.

The mean number of communication items per problem for each position is shown in Table 5. Analyses of variance for each net taken separately showed that in the slash positions differed significantly ($p < .001$), and in the wheel positions and the distribution × positions interaction were significant ($p < .001$). Examination of the means in Table 5 shows that in all nets the individual centrality indices were related to the number of items communicated, with those positions having the higher individual centrality sending more messages. In the wheel and the slash, the person receiving the most information sent more items than did the equivalent position under conditions of equal information distribution. Again, an inverse relationship was found between the time required to reach a solution and the number of items transmitted.

Errors. Two types of errors were computed: (*a*) decision errors, defined as an incorrect answer reported by a member of the group, and (*b*) information errors, defined as the transmission of an item of information different from that contained in, or derivable from, the items of information given to the group as a part of the problem to be solved. Since the error scores were decidedly skewed and the variance was heterogeneous, the Mann-Whitney *U* test (6) was used instead of the usual analysis of variance. Only the differences produced by the information distribution variable were significant ($p < .001$). The mean number of decision errors was 2.6 and 2.1 for equal and unequal distribution, respectively; the mean number of information errors was 9.8 and 6.2 for equal and unequal distribution, respectively. This means that unequal information distribution decreased the number of both decision and information errors.

The corrective power of the nets as measured by the ratio of decision errors to information errors is given in Table 6. Analysis of variance showed significant differences between nets ($p < .05$), but information distribution had no effect. The gap test showed that the wheel differed significantly from the other two nets; i.e., the wheel was less effective in this respect, since a larger numerical value indicates less corrective power. This finding is in agreement with the findings with respect to time scores and number of items transmitted: the wheel differs from the other nets but there is no precise relationship between corrective power and net centrality.

Questionnaire results. The data obtained from the questions requiring sociometric choices of the *S* most preferred in various situations were analyzed for each net taken separately. In the wheel, 36 chi squares were computed to test this hypothesis and 28 of these were significant ($p < .05$). In the slash, 19 chi squares were significant ($p < .05$) out of 24 computed. In the circle, only 4 chi squares were significant ($p < .05$) out of 48 computed. These results offer strong support for the hypothesis that *S*s in the positions having the higher centralities are chosen more frequently than *S*s in positions having lower centralities.

Comparisons of choices under the two conditions of information distribution yielded only

TABLE 6

CORRECTIVE POWER OF NETS: RATIO OF DECISION ERRORS TO INFORMATION ERRORS*

Measure	Equal Distribution			Unequal Distribution		
	Wheel	Slash	Circle	Wheel	Slash	Circle
Mean	.51	.33	.30	.77	.34	.25
SD	.59	.39	.33	.83	.37	.41

* Groups making no information errors were omitted when computing corrective power.

nonsignificant chi squares. The *S* in the position receiving the most information was chosen more frequently under conditions of unequal distribution of information than was the equivalent position under conditions of equal distribution of information in 16 of the 18 comparisons made, but the differences were small and hence did not result in a significant change in the over-all distribution of choices.

Examination of the sociometric choices revealed a similar pattern with respect to the number of times a single *S* was chosen unanimously as the one most preferred. Under conditions of equal information distribution, the mean number of unanimous choices was 6.5, 3.2, and 0.8 for the wheel, the slash, and the circle, respectively; under conditions of unequal information distribution, the corresponding means were 6.0, 1.5, and 2.3. The nonparametric test developed by Friedman (3) and described by Wilcoxon (8, pp. 7–9) was used to analyze these data. Over-all differences were highly significant ($p < .001$), but when net and distribution effects were isolated, only the net differences were significant ($p < .01$). However, the net × distribution interaction was significant ($p < .01$), indicating the effects of the net were different for different distributions.

Analyses of both the number of times chosen and the number of unanimous choices indicate that emergence of a leader, as indicated by sociometric choice, is directly related to the centrality index when only the equal information distribution situation is considered. In the unequal distribution of information situation, the number of unanimous choices was increased in the circle, decreased in the slash, and essentially unchanged in the wheel. This finding indicates that emergence of leadership is determined by the number of positions in the net having the highest individual centrality (inverse relationship) and the maximum difference in individual centralities (direct relationship).

The mean ratings per net for the questions dealing with job satisfaction, positions comparison, group cooperation, and group performance were analyzed by applying analysis of variance to each question taken separately. The results may be summarized as follows: (*a*) *S*s in the wheel enjoyed their job in the group significantly less than did the *S*s in the slash and in the circle ($p < .001$). (*b*) As a group, the *S*s showed no significant differences in their ratings of the relative value of positions in the net. (*c*) The ratings of group cooperation showed slight but significant net differences ($p < .05$); the *S*s in the wheel rated the cooperation of their group lower than did the *S*s in the slash who in turn rated the cooperation of their group lower than did the *S*s in the circle. (*d*) The *S*s in the wheel rated the performance of their group lower than did the *S*s in the slash and the circle under all conditions; under conditions of equal distribution of information, the slash and the circle did not differ in this respect, but under conditions of unequal information distribution, the *S*s in the slash rated the performance of their group higher than did the *S*s in the circle ($p < .05$).

Taken together, the rating scale results indicate that under conditions of equal information distribution, centrality was not clearly related to group morale (where morale is defined as general satisfaction with the group's activity). Unequal distribution of information affected morale in only one of the three nets; it seems safe to conclude that information distribution is not an effective variable as far as group morale is concerned. These results are in agreement with the findings with respect to time scores, communication units, and corrective power in that the wheel differed from the other two nets and in that the information distribution variable had little effect upon the group scores.

The rating scale data were also analyzed according to the ratings of individual positions, with the following results: (*a*) There were no significant differences between the ratings of job satisfaction and group cooperation. (*b*) The ratings of various positions differed significantly ($p < .01$) under all experimental conditions, and the distribution × positions interaction was significant ($p < .05$) for the wheel and the circle. The positions having higher centrality indices were rated higher than those positions having lower centrality indices under both conditions of information distribution and in all three nets. In the wheel and the circle, the position receiving most of the information was rated higher under conditions of unequal distribution than was the equivalent position under

conditions of equal information distribution. (*c*) In the slash, all positions rated the group performance higher under conditions of unequal information distribution than under conditions of equal information distribution ($p < .001$).

These results lead to the conclusion that there is some relationship between individual centrality and individual morale, but this relationship is tied very closely to the position preference component. The information distribution variable produced equivocal results: additional information either had no effect upon individual morale (as indicated by job satisfaction and group cooperation ratings) or increased the individual morale (as shown by the positions comparisons ratings in the wheel and the circle and the group performance ratings in the slash).

Responses to the question concerning group organization revealed that all groups used one of two forms of organization: (*a*) all information was sent to one *S* who solved the problem and sent the answer out to the other *S*s, or (*b*) all information was sent to all *S*s and then each *S* solved the problem independently. The first method was used by only one group in the slash and by only one group in the circle, but was used by eleven groups in the wheel. These data do not support the earlier findings that high net centrality leads to more consistent group organization.

DISCUSSION

The data from this study make it clear that centrality correlates with behavioral measures only when differences among positions within the net are considered. Similarly, altering the distribution of information changes the behavior of *S*s in the various positions, but has little effect upon the performance of the group.

Let us examine some of the positive relationships among the variables in the present study. Subject to the limitations imposed by the conditions under which the data were collected, the results of this experiment offer some evidence for the following hypotheses:

1. The time required for an individual in a position to complete an activity varies inversely with individual centrality.

2. Individual morale, number of items transmitted by an individual in a position, and the probability that an individual in a given position will be chosen as a leader vary directly with individual centrality.

3. The emergence of a leader in a net (as indicated by unanimous choice) is determined by the number of positions having the highest individual centrality index in that net and the maximum difference in individual centralities.

4. Increasing or decreasing the amount of information initially assigned to an individual in a given position has an effect similar to respectively increasing or decreasing the individual centrality index of that position.

How can these relationships be explained? Leavitt believed that centrality was effective primarily because it determines the degree of independence of the individual in a position. The present study offers no evidence against this explanation, but there is another possible explanation of the relationship between individual centrality and morale that perhaps should be mentioned. This view would assume that centrality, because of its correlation with number of channels,[7] determines the amount of communication activity engaged in by an individual and hence his enjoyment of the task. This draws some support from an *ad hoc* analysis of time per message which showed that *S*s with higher centralities spent significantly less time per message than did *S*s with lower individual centralities ($p < .001$). The implicit assumption is that an individual enjoys a task which keeps him occupied more than he does a task which permits relatively long periods of inactivity.

The reason for the relationship between individual centrality and number of items transmitted becomes obvious when it is remembered that individual centrality and number of channels are perfectly correlated and that the most frequent method of solution was for each *S* to send all available information to all other *S*s with whom he was connected. Thus it was inevitable that more channels would require more messages.

It has been pointed out that the emergence of leadership in a net is determined by the number of positions having the highest centrality index in the net and the maximum

[7] Centrality and number of channels are perfectly correlated in the nets used in this experiment, but this is not true for all possible nets.

difference in individual centralities. The interpretation would seem to be that the less ambiguity associated with the distinction between the position having the highest individual centrality and the other positions, the more agreement there is among group members regarding which *S* is most desirable for the leadership role.

The effect of the communication net upon corrective power requires some explanation. It appears that in the wheel the only ways in which an error could be detected would be for an *S* to detect his own error, for an *S* to pass on to another *S* a set of computations containing an error, or for an *S* to pass two conflicting statements to the same *S*. In the slash and in the circle an additional opportunity for error detection was possible: an *S* might receive information about the same objects or events from two or more different sources. This added opportunity for error detection and correction probably accounts for the superiority of the slash and the circle in this respect.

In summary, it is believed that individual centrality determines the behavior of individuals in various positions within nets via the limitations which it imposes on these individuals. Centrality, however, bears no simple relationship to independence of action. The underlying processes which determine behavior vary with the particular aspect of behavior which is being measured.

SUMMARY

The purposes of the present experiment were (*a*) to test the effects of unequal distribution of information upon leadership, morale, and problem-solution efficiency in various communication nets, and (*b*) to check the findings of previous investigators, using a set of slightly more complex problems and different communication nets. It was believed that the primary source of differential forces would be centrality, and that increasing or decreasing the amount of information initially assigned to a given position would be equivalent to respectively increasing or decreasing the centrality index of that position.

Fifteen groups were randomly assigned to each of six experimental conditions, representing the six possible combinations of two types of information distribution (equal and unequal) and three communication nets (the wheel, the slash, and the circle). Each group was required to solve three eight-item problems.

The results were interpreted as supplying some evidence for the following hypotheses:

1. Time required for an individual in a position to complete an activity varies inversely with individual centrality.

2. Individual morale, number of items transmitted by an individual in a position, and the probability that an individual in that position will be chosen as a leader vary directly with individual centrality.

3. The emergence of leadership in a net varies inversely with the number of positions having the highest individual centrality index in that net and directly with the maximum difference in individual centralities.

4. Increasing or decreasing the amount of information initially available to an individual in a given position has an effect similar to respectively increasing or decreasing the individual centrality index of that position or the number of channels available to that position.

REFERENCES

1. BAVELAS, A. A mathematical model for group structure. *Appl. Anthropol.*, 1948, **7**, 16–30.
2. BAVELAS, A. Communication patterns in task-oriented groups. *J. acoust. Soc. Amer.*, 1950, **22**, 725–730.
3. FRIEDMAN, M. Use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Amer. statist. Ass.*, 1937, **32**, 675–701.
4. HEISE, G. A., & MILLER, G. A. Problem-solving by small groups using various communication nets. *J. abnorm. soc. Psychol.*, 1951, **46**, 327–335.
5. LEAVITT, H. J. Some effects of certain communication patterns on group performance. *J. abnorm. soc. Psychol.*, 1951, **46**, 38–50.
6. MANN, H. B., & WHITNEY, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, **18**, 50–60.
7. TUKEY, J. W. Comparing individual means in the analysis of variance. *Biometrics*, 1949, **5**, 99–114.
8. WILCOXON, F. *Some rapid approximate statistical procedures*. Stamford, Conn.: American Cyanamid Co., 1949.

Received July 9, 1953.

SOME EFFECTS OF UNEQUAL DISTRIBUTION OF INFORMATION IN A WHEEL GROUP STRUCTURE[1]

J. C. GILCHRIST, MARVIN E. SHAW, AND L. C. WALKER

University of Wisconsin

IN THE immediately preceding report, Shaw (1) investigated some effects of unequal distribution of problem-related information in group structures. In general, he found that information distribution within structures had no effect upon over-all structure measures of time, message units, errors, or morale although decreasing or increasing the amount of information initially available to an *individual* within a structure had an effect similar to respectively decreasing or increasing the individual centrality index of that position.

There are, however, other distributions which affect the generality of this finding. In any structure having central person(s) information may be distributed so that a peripheral person receives more or so that a central person(s) receives more. Shaw's experiment tested only the former possibility: the present experiment tests both possibilities using the four-person wheel structure.

PROCEDURE

The apparatus, problems, questionnaires, and procedures were identical to those of Shaw.

The *S*s were students enrolled in psychology and/or sociology courses at the University of Wisconsin. Forty-five groups of four members each were run. Thirty of these groups were composed of females and fifteen were composed of males. Each of these groups was randomly assigned to one of three experimental conditions, subject to the restriction that each experimental condition be assigned ten groups of females and five groups of males. The three experimental conditions were: (*a*) equal distribution, (*b*) unequal distribution to the periphery, and (*c*) unequal distribution to the center.

Each group was required to solve three problems randomly assigned one at a time. Under the condition of equal information distribution, for each problem each *S* was given a card stating the problem and two cards giving information required in the problem's solution. Under conditions of unequal distribution, the *S* in the appropriate position (A in the peripheral condition and B in the central condition. See the "wheel" in Fig. 1 of Shaw's report) was given one of the problem cards and five of the information cards, and each of the other *S*s was given one of the problem cards and one of the information cards.

RESULTS

A simple within and between analysis of variance performed upon the times for the single fastest correct trial yielded an insignificant value of F.

Analyses of variance were performed upon the times of the fastest position per trial and slowest position per trial. Distribution was found to have no effect, either main or interaction. The significant effects in both analyses were trials and groups within. The former indicated that learning was occurring; the latter that the sets of four persons differed from each other. (The groups within term is equivalent to an individual differences term in the usual psychological design.)

A further breakdown of time scores involves positions within the structures. The time measures here for the main effects of distribution and trials and their interactions differ from the previous two analyses in that here the design results in mean time appearing as means for those variables, i.e., the mean of the time of completion of the four positions' times. Trials are again significant, and show a learning effect. Positions and distribution × positions are also significant in addition to individual differences terms. Because position when tested by its interaction with distribution is not significant, only the interaction means are given in Table 1. Note particularly that the central position is affected adversely as to time under unequal central distribution.

The measure of communication was the message unit defined as a simple sentence or a meaningful part of a compound or complex sentence. The significant terms of the analysis of variance are discussed below. Trials show, once again, a learning effect, i.e., fewer message units per trial. Position, remaining significant over and above the distribution × position interaction, shows that the central person sent 3.3 times as many message units as the other positions. Table 2 summarizes the means of the distribution × position interaction.

[1] This research was supported in whole or in part by the United States Air Force under Contract AF 18(600)-340, monitored by the Human Resources Research Institute, Maxwell Air Force Base, Alabama.

TABLE 1

MEANS AND RESULTS OF TUKEY'S GAP TEST (2) OF TIME SCORES BROKEN DOWN BY POSITIONS

Position	Information Distribution		
	Equal	Unequal Peripheral	Unequal Central
A	11.04	9.05*	11.14
B	9.74*	10.91	10.54
C & D	11.18	11.13	11.37

* Separated as a straggler ($p < .05$).

TABLE 2

MEANS AND RESULTS OF TUKEY'S GAP TEST OF THE DISTRIBUTION × POSITION INTERACTION OF NUMBER OF MESSAGE UNITS TRANSMITTED

Position	Information Distribution		
	Equal	Unequal Peripheral	Unequal Central
A	9.69	12.24†	7.31
B	29.60*	27.00*†	31.82*
C & D	10.11	7.88	7.73

* Separated by a significant gap ($p < .05$) as a homogeneous group.

† Separated as a straggler ($p < .05$).

Errors were analyzed as in Shaw's experiment. No significant differences were found on any of the error measures.

Analysis of the sociometric questions using chi square demonstrates that information distribution did not change the pattern of sociometric choices on the six questions. Of the 24 chi squares computed between distributions, only one was significant. These chi squares were computed for (*a*) the A position choosing, (*b*) the B position choosing, (*c*) the C and D positions combined choosing, and (*d*) all positions combined choosing for each question.

On the other hand, chi squares computed within information distribution groups show that the choices reflect the group structure: Positions A and C plus D chose B ($p < .05$ for 25 of 36 chi squares). Position B chose randomly among A, C, and D.

The rating-scale data were subjected to analyses of variance. It was found that position was effective on three of the four questions. In each instance, the B position yields the high mean, i.e., is more satisfied with the job, considers its status higher, and believes the group cooperated better than the other three positions. Of most interest, however, are the significant interactions of distribution × position on the first two questions. The means of these interactions are given in Table 3. Note that every rating change follows the connected propositions that individual morale is directly related to individual centrality and that increased or decreased information status is similar to respectively increased or decreased individual centrality.

TABLE 3

MEANS OF DISTRIBUTION × POSITION INTERACTION ON JOB SATISFACTION AND POSITION STATUS RATINGS*

Position	Job Satisfaction			Position Status		
	Equal	Unequal Peripheral	Unequal Central	Equal	Unequal Peripheral	Unequal Central
A	5.2	6.5	5.2	10.4	12.9	10.8
B	6.2	5.9	6.8	17.5	16.2	18.9
C & D	5.4	4.2	4.9	11.7	9.0	10.3

* Tukey's gap test yields no more information in this instance than does the analysis of variance.

Categorizing the answers to a question regarding the groups' development of a system showed no differences between the information distribution conditions when tested by chi square. Sixty-seven per cent of the groups, however, adopted the system of sending all information to everyone and checking each other's answers. In the remaining 28 per cent which developed a system, the central position, B, received all information and transmitted only answers. Two groups claimed that no system was developed.

DISCUSSION

The data show that individual centrality is (*a*) directly proportional to individual morale and to the probability that an individual will be chosen as a leader and (*b*) inversely proportional to individual problem-solution times where there is equal distribution of information, but that there are deviations under other information-distribution conditions. Since the argument will be based primarily upon individual-position time scores, they are discussed first.

Table 1, it will be recalled, presented the means of the position × information-distribution time scores. Shaw's formulation that increasing or decreasing information at a position is similar to respectively increasing

or decreasing the centrality index of that position is upheld there as long as comparisons are made *within* distribution conditions.[2] Once comparisons are made between distributions, however, failure to follow Shaw's proposition is apparent. The most noticeable failure is at position B in the unequal central distribution which should be the lowest mean and, instead, is the third lowest. This negative instance must be accounted for, and a probable answer is found in the communication requirements placed upon B by the nature of the information distribution.

The groups usually adopted the system of "information to everyone." Under this system, B must send a minimum of 17, 18, and 21 message units in the unequal peripheral, equal, and unequal central conditions respectively. As seen in Table 2, the message units actually sent by B order directly in the sequence given by these minima.[3] Times of the B position, though, order equal, unequal central, unequal peripheral. These orderings suggest that there is an interaction between individual centrality and communication requirements: that, when the number of required message units passes a certain optimal output level, it begins to counteract the effects of individual centrality. Such a statement may best be summarized with the term "saturation"[4] since it is not simply and directly related to the number of message units required to be transmitted as is shown by the A position in the unequal peripheral distribution. Thus, information distribution is indirectly an effective variable by virtue of its effects upon what is, perhaps, a more basic phenomenon, saturation, or the input and output requirements which it places upon positions within a group structure.

One other comment is pertinent. Although morale follows almost perfectly Shaw's original statement as to the effects of information distribution, the emergence of leadership does not: information distribution did not affect leadership emergence. Examination of Shaw's data indicates that that conclusion was based upon the findings in structures where each position has two or more channels terminating upon it. Information distribution, then, has an effect upon leadership emergence in structures having lesser differences in centrality than does the wheel structure.

SUMMARY AND CONCLUSIONS

An experiment was conducted using a four-person wheel group structure to test the effect of information distribution upon problem solving, leadership, and morale within the structure. Three conditions of information distribution were used: equal, unequal with most information initially located at the center of the structure, and unequal with a peripheral position initially receiving the most information.

The data substantiate the position that the individual centrality index is directly related to the emergence of leadership in the structure and individual morale. The findings with respect to information distribution, however, suggest that it is not simply and directly related to the individual centrality index. Rather, it appears to be only indirectly related thereto and mediated by a more basic phenomenon which has been termed "saturation."

[2] There is some question as to the comparative time of positions A and B in the unequal peripheral condition. Should this distribution so readily overcome the structural centrality effect as to produce the time differences between those positions that it did? The difference between these means is associated with a $p < .05$.

[3] The differences between the means are almost identical with the differences between the minima. This suggests, since the absolute values of the means are higher than those of the minima, that the channels carry a constant amount of "noisy" and/or redundant messages. Examination of Shaw's data for three nets shows the same effect appearing there.

[4] Two kinds of what we have called saturation may be distinguished: one we call "channel saturation" and refers to the number of channels with which a position must deal; the other, discussed above, is "message unit saturation." The former should be closely related to both centrality and group size and the latter to centrality, problem size, and information distribution. Each of the two main classes may break down into "input" and "output" saturations. If these variables are basic ones and the above textual statements are correct, they should ultimately affect leadership emergence and morale. Experiments now in progress bear upon these points.

REFERENCES

1. SHAW, M. E. Some effects of unequal distribution of information upon group performance in various communication nets. *J. abnorm. soc. Psychol.*, 1954, **49**, 547–553.
2. TUKEY, J. W. Comparing individual means in the analysis of variance. *Biometrics*, 1949, **5**, 99–114.

Received August 25, 1953.

THE EFFECT OF ONE PARTNER'S SUCCESS IN A RELEVANT TASK ON THE INTERACTION OF OBSERVER PAIRS[1]

BERNARD MAUSNER

University of Massachusetts

THE PRESENT report is based on one of a series of studies of the determinants of social interaction in judgment. The paradigm for these investigations is that of the Sherif experiment in which it was demonstrated that the judgments of subjects (*S*s) reporting the extent of autokinetic effect converged toward a group norm. It has already been found that the degree of this convergence can be affected by such factors as prior reinforcement of the responses (2, 4, 7), variation in the psychophysical method employed (1, 5), variation in the nature of the stimuli (3), and manipulation of the prestige of one of the partners through his identification as an "expert" or "nonexpert" (6). With reference to the last of these factors, there exists the problem of specifying in as precise a manner as possible the variables determining prestige. When this is done, prestige may serve as a useful intervening variable in the theoretical treatment of social interaction.

The writer (8) has suggested that convergence in judgment is a function of the summation of two conflicting response tendencies: (*a*) a tendency to continue to make previous judgments, and (*b*) a tendency to agree with a partner. The partner's prestige is probably one of the determinants of the latter tendency. Prestige may be strengthened where the partner gives specific evidence of competence in the kind of judgment being made; it may be weakened where evidence of incompetence is presented. Other things being equal, this variation in prestige should result in a concurrent strengthening or weakening of the tendency to agree with the partner. Where the opposing tendency, that to maintain an earlier range of judgments, is of known strength, or is judged to be randomized in a group of *S*s, convergence should be predictable from information concerning the partner's demonstrations of competence.

On the basis of these considerations a specific hypothesis was tested: *Ss will show a greater tendency to converge in judgment toward partners who have previously demonstrated success in a related task than toward partners who have previously demonstrated failure.*

METHOD

Subjects. The *S*s were 28 undergraduate students in elementary psychology classes at Washington Square College, New York University. All were naive with respect to the purposes of the experiment and the area of the investigation.

Procedure. In a preliminary session each *S* judged the length, in inches, of each of a series of 100 lines. These consisted of ten lines varying in length from 10 to 19 in., each presented ten times. Each block of ten trials contained all ten lengths arranged in a different permutation from that used in other blocks. The judgments were made in a dark room; the lines were presented by means of an apparatus in which a lighted slot ½ in. wide and varying in length was exposed to *S* for approximately 1 sec. (for complete description cf. 4).

Following this session (the Alone session), a confederate of *E* was presented as a second "subject." Both *S*s were given 20 trials in which they were instructed to move a stylus through a 17-cm. slot placed behind a curtain. They then attempted to reproduce this length by moving the stylus through an equivalent distance in a free slot, also hidden. The actual *S* completed the task first and was followed by the confederate. Each *S* was requested to observe and record his partner's performance. A success was defined as reproduction within 1 cm. of the correct distance. The confederates were instructed to perform very skillfully for some *S*s, very unskillfully for the remainder. Those *S*s for whom the partner's reproductions had an *SD* of 1.5 cm. or better were considered to have had a "successful" partner (group S); those for whom the *SD* of the reproductions was more than 1.5 cm., an "unsuccessful" partner (group U). The *S*s were assigned to these groups in rotation. Group S contained 17 *S*s, group U, 11 *S*s.

After this procedure had been completed, both *S*s judged the lengths of the same series of lines previously presented (Together series). They alternated as first to respond to each stimulus. The confederates had been instructed by *E* to make their judgments vary consistently by 4 to 5 in. from those given by the actual *S* in the alone situation.

Treatment of results. Since one of the two persons in each pair was *E*'s confederate, and, by instruction, was

[1] The present report is the fourth in a series entitled, "Studies in Social Interaction." The data were presented at the 1953 meetings of the Eastern Psychological Association. The experiment was carried out in the laboratory of the department of psychology, Washington Square College, New York University. The assistance of the following students who served both as confederates and as experimenters is gratefully acknowledged: Leon Dreier, Peter Sciaki, Marilyn Reisman, Sandra Skolnick, and Barbara Diamond.

forced to maintain a relatively constant range of judgments, the degree of social interaction in judgment was estimated by measuring the degree of shift in judgments made by the other member of the pair, the actual *S*, in two ways. First, a D score was calculated for each *S*. This measure is derived from the difference in frequency with which each number response was made in the alone and together situations. A high D score indicates considerable shift, and, where this is toward the partner, considerable convergence. A low D score indicates maintenance of the same range of judgments in the together as in the alone session (see 4, 6, for detailed method of calculation). In the second method of analysis, graphs of median judgments per block of ten trials were used to show direction of shift and as a basis for qualitative descriptions of the progress of interaction through the course of the experiment.

RESULTS AND DISCUSSION

Table 1 gives D scores for the *S*s of both groups. The mean D score for group S is 62.1, for group U, 20.2. This difference is significant at the .01 level ($t = 3.10$, $n = 26$, $p < .01$). This indicates greater shift in judgment from alone to together situations for the *S*s whose partners were successful (group S) than for those whose partners were unsuccessful (group U). Inspection of the median graphs shows a tendency to formation of a social norm of judgment in the pairs of group S, a lack of such tendency in the pairs of group U. Figure 1 shows three such plots for *S*s in group U; Fig. 2 four plots for *S*s in group S. These are representative of the 10 *S*s in group U and the 12 in group S whose behavior corresponded with the prediction.[2] Thus the hypothesis presented above is substantiated by the data of this experiment.

Six of the 28 *S*s do not show the predicted behavior. Four of the *S*s in group S have D scores below the mean for group U; one (SWI) is borderline. These are clearly distinguished from the remainder of the *S*s in this group both by their D scores, and by the failure of their graphs to show formation of a social norm. The *S*s' own performance on the interpolated task was not at all related to the degree of social interaction. Among the five *S*s in group S who failed to converge none was particularly accurate in "blind" slot reproduction. The two most accurate *S*s have the highest and the third highest D scores. The most inaccurate *S* in the entire experimental group (FOT) is found in group U, and has an appropriately low D score (12); i.e., he failed to converge.

Unfortunately, there is no further information available in the data of this experiment to throw light upon the behavior of these deviant *S*s. However, their occurrence may be predicted from the theoretical scheme alluded to above. In any group of *S*s chosen at random it is likely that, owing to personality factors or other variables not immediately identifiable, some will have a high tendency to maintain previous judgments. The opposing tendency to agree with the partner is probably not too

TABLE 1

INTERACTION BETWEEN *S* AND *E*'S CONFEDERATE (*P*) GIVEN BY D SCORE; ACCURACY OF SLOT REPRODUCTION BY *SD* OF TRIALS; GROUP S: PARTNER'S *SD* < 1.50, GROUP U: PARTNER'S *SD* > 1.50

Subject	D	Slot Reproduction	
		SD(*S*)	*SD*(*P*)
Group S			
ZVI	67	3.17	.38
KLE	58	1.51	.70
HEN	78	2.21	1.39
KEL	84	.80	.90
BRE	68	1.84	.83
LAN	55	2.10	.84
HAM	110	2.42	1.41
DEB	60	1.73	1.39
NAS	68	2.64	1.05
HOO	123	.94	1.24
BRA	162	2.88	.83
FRI	60	2.93	.78
SWI*	30	2.35	.92
COH*	4	2.00	.57
TAN*	10	2.73	1.17
HEI*	6	2.40	.77
CHA*	13	2.63	1.39
Group U			
SPO	7	2.26	3.36
THE	6	3.78	2.28
AUE	2	2.28	3.12
BAS	24	.40	2.17
CAL	19	2.95	1.88
FOT	12	7.10	2.93
MER	10	3.20	2.41
MAS	16	3.60	4.10
JUS	31	1.75	1.84
ECK	32	2.24	1.58
HOL*	66	3.43	2.87

* *S*s whose results did not conform to the prediction.

[2] Graphs of median judgments per block of ten trials for the remainder of the *S*s have been deposited with the American Documentation Institute. Order Document No. 4113 from the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for 35 mm. microfilm, or $1.25 for 6 × 8 in. photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

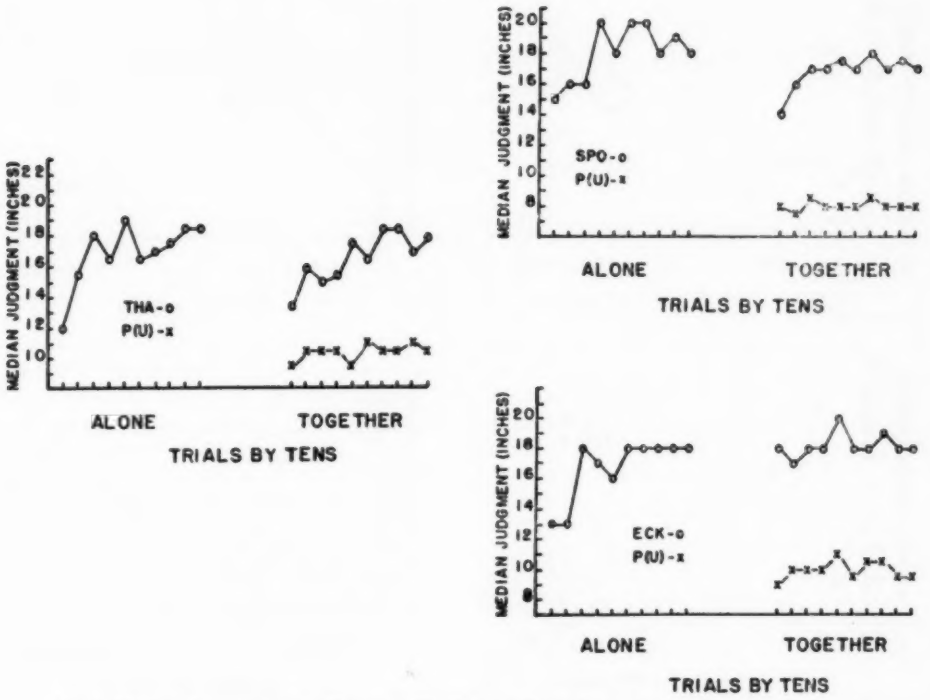


FIG. 1. MEDIAN JUDGMENTS PER BLOCK OF TEN TRIALS FIRST ALONE AND THEN WITH A PARTNER UNSUCCESSFUL IN SLOT REPRODUCTION. PLOTS FOR THREE *Ss* ARE INCLUDED.

high to begin with where the partner is an anonymous fellow student. It would be unlikely, therefore, that a manipulation of the partner's prestige, such as that here attempted, would be effective 100 per cent of the time. The fact that it was effective in 71 per cent of the cases in group S, and that a significant difference in D scores was found between the two groups, argues for the sensitivity of the convergence experiment as a measure of social interaction.

Apparently the effect of a partner's failure in lowering the tendency to agree is even more pronounced than the effect of his success in raising it, since all but one of the *Ss* in group U failed to converge. That deviant case (HOL) had a D score (66) actually above the mean for group S. Although there is no corroborating evidence, it is possible that the fact that HOL is the sole Negro *S* is significant. There may be factors in the developmental history of Negro *Ss* which would lead to a heightened tendency to agree with a white partner.

Thus two kinds of conditions antecedent to prestige can now be identified. Prestige as measured in terms of convergence in judgment has been anchored to a partner's group membership (7) and may now also be anchored to a partner's demonstrations of competence in relevant tasks. The next step in these investigations should be an attempt to identify and measure some of the genetic factors which relate prestige to tendency to agree. It may be suggested that, for most people, agreement with prestigeful individuals has been positively reinforced, and disagreement negatively reinforced. That is, most *Ss* used in these experiments have received reinforcements from agreeing with members of an "expert" group, or with people who demonstrate competence. This argument leads to a testable hypothesis. It may be predicted that *Ss* will converge toward partners with whom *agreement* has previously been positively reinforced in some other context, but will not converge toward partners with whom agreement has been negatively reinforced. If such a relation-

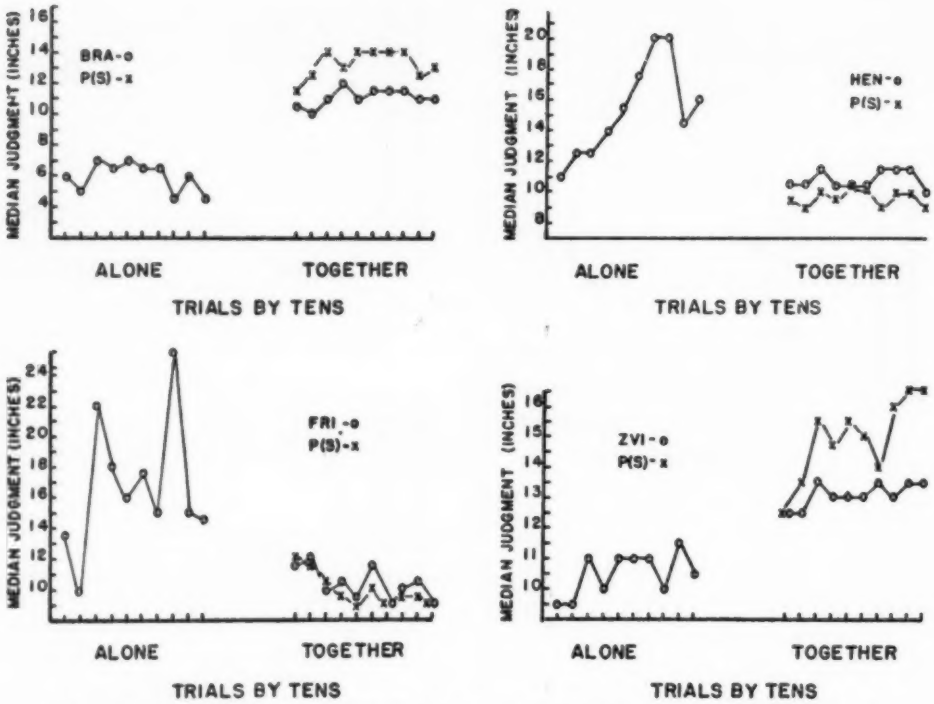


FIG. 2. MEDIAN JUDGMENTS PER BLOCK OF TEN TRIALS FIRST ALONE AND THEN WITH A PARTNER SUCCESSFUL IN SLOT REPRODUCTION. PLOTS FOR FOUR *S*s ARE INCLUDED.

ship were found to hold, the kinds of prestige effects demonstrated in the present and earlier studies could logically be related to similar histories of differential reinforcement in our *S*s' development.

SUMMARY AND CONCLUSIONS

Twenty-eight undergraduate *S*s judged the lengths of 100 lines first alone and then with a confederate of *E*. These confederates had demonstrated failure in a similar task for 11 of the *S*s, and had demonstrated success for 17 of them.

The *S*s who worked with previously successful partners converged significantly more toward these partners in the group judgment situation than did the *S*s who worked with previously unsuccessful partners.

This finding is held to indicate that demonstrations of success and failure in relevant tasks may be used as one kind of antecedent condition anchoring the intervening variable "prestige." A partner's prestige is considered one of the determinants of the tendency of an *S* to agree with him, and thus finally of the degree of convergence in judgment.

REFERENCES

1. BERENDA, RUTH W. *The influence of the group on the judgments of children*. New York: King's Crown Press, 1950.
2. KELMAN, H. Effects of success and failure on "suggestibility" in the autokinetic situation. *J. abnorm. soc. Psychol.*, 1950, **45**, 267–285.
3. LUCHINS, A. S. On agreement with another's judgments. *J. abnorm. soc. Psychol.*, 1944, **39**, 97–111.
4. MAUSNER, B. The effect of previous reinforcement on the interaction of observer pairs. Univer. Microfilm, Ann Arbor, Mich., Publ. No. 2118, 1950.
5. MAUSNER, B. The effect of variation in psychophysical methods on the interaction of observer pairs. Paper read at Eastern Psychol. Ass., Atlantic City, N. J., April, 1952.
6. MAUSNER, B. Studies in social interaction: III. The effect of variation in one partner's prestige on the interaction of observer pairs. *J. appl. Psychol.*, 1953, **37**, 391–393.
7. MAUSNER, B. The effect of prior reinforcement on the interaction of observer pairs. *J. abnorm. soc. Psychol.*, 1954, **49**, 65–68.
8. MAUSNER, B. Studies in social interaction: I. A conceptual scheme. *J. soc. Psychol.*, in press.

Received July 15, 1953.

SOME PERSONALITY CORRELATES OF STIMULUS GENERALIZATION UNDER STRESS

CHARLES W. ERIKSEN

The Johns Hopkins University

IN THEIR recent review of research on psychological stress, Lazarus, Deese, and Osler (11) have pointed out the importance of the individual differences component in performance under stress conditions. In some subjects (*S*s) the introduction of stress results in poorer performance, but in others performance is unaffected or even improves. The purpose of the present experiment was to investigate some of the personality correlates of these individual differences in response to stress. Specifically, the problem here was the relation of the hysteria-psychasthenia dimension, as measured by the respective scales in the Minnesota Multiphasic Personality Inventory (MMPI), to the degree of stimulus generalization shown under the stress of strong electric shocks. The *S*s were trained to make a simple horizontal arm movement when a square of a certain size was presented, and the amount of generalization of this response to squares of other sizes was determined.

Previous research has indicated that learning and memory performance under stress produced by "self-esteem" threat is related to the hysteria-psychasthenia variable. The *S*s scoring high on hysteria have been found to have poorer recall of failures on self-esteem–involving tasks (6, 15) and poorer memory for anxiety-arousing stimuli (1) than *S*s scoring high on psychasthenia. However, in the present experiment not only was a different aspect of performance measured, but the method of inducing stress was the threat of painful physical stimulation rather than a threat to the *S*s' self-esteem. To see what relationship, if any, might be expected between performance and the hysteria-psychasthenia variable under these changed conditions, we need to consider briefly some of the factors involved in the concept of psychological stress.

We can equate psychological stress with anxiety and analyze it into the following sequential components:

Stimulus situation → anxiety response → *S*'s response to his internal anxiety-produced stimuli

Such an analysis helps us to bear in mind some of the factors involved in stress and to see what effect changes in the stress conditions will have upon different performance measures.

The stimulus situation alone provides ample opportunities for individual differences to occur in stress research. It has been shown (12) that differences in goals and prior experiences may render a given stimulus anxiety arousing for some *S*s and not for others. More pertinent to our present problem is how differences in the stimulus situation limit or affect *S*'s response to his internal anxiety cues. Responses are made not only to the internal anxiety cues, but they must also be adapted to the external anxiety-provoking situation.

Although a number of different stimulus situations have been used to stress *S*s, most, if not all, represent variations upon one or the other of two basic types. One is to threaten *S*s with painful physical stimulation such as electric shocks. The other is to threaten self-esteem by failure or the threat of failure on some highly competitive task. Both types of situation and their variations have been successful in making *S*s anxious, but it is to be noted that a given response may not be equally effective to both types of threat. People differ in the defenses they use and in the way they have learned to handle and interpret anxiety. We may expect that some of these learned responses are more appropriate to one type of stress than to the other. Avoidance defenses such as forgetting failures or "psychologically leaving the field" may be effective methods of evading anxiety under situations of self-esteem threat, but they are quite useless when it comes to evading electric shocks or other physical punishment.

In the present experiment, where anxiety was induced by electric shocks, the prediction was made that *S*s scoring high on the hysteria scale would show more stimulus generalization than *S*s scoring high on psychasthenia. This prediction was arrived at from two considerations. First, *S*s high on hysteria, deprived of using their avoidance-type of defense by the nature of the present stress situation,

should be more subject to uncontrolled anxiety. The *S*s high on psychasthenia, on the other hand, would appear from previous research and from the content of the MMPI items to be more accustomed to coming to grips with anxiety on the conscious level. Their supposed defenses of rationalization and intellectualization should be more compatible with the present stress situation.

The second consideration was that the greater the uncontrolled anxiety, the greater the stimulus generalization. Meyer (13) has recently summarized evidence indicating that one effect of anxiety is to lower the threshold for responses, and experiments by Hilgard, Jones, and Kaplan (10) and by Rosenbaum (14) have suggested a relation between anxiety and stimulus generalization.

METHOD

Design. The *S*s were trained to respond with a horizontal arm movement when a square of a certain size (training stimulus) was presented. After this response had been well established, squares of different sizes (generalization stimuli) were occasionally presented to *S*s instead of the training stimulus, and *S*s' responses or tendencies toward responses were recorded. Two experimental conditions were used. Under the *avoidance condition*, *S*s were instructed that on some of the trials when the training stimulus appeared a shock circuit would be *on*. On these trials they would receive a shock if their response to the training stimulus was not rapid enough. Under the *nonavoidance condition*, *S*s were also told that some of the time when the training stimulus appeared a shock circuit would be *on*, but here they were given to understand that the shocks were completely unrelated to their performance. In other words, no matter what they did, when the shock circuit was on, they would be shocked.

These two experimental conditions were used to determine what effect the factor of evitability of shock might have upon performance and also as a control for *S*s who might deliberately decide to respond to every stimulus in order better to avoid receiving shocks. As a further control against this latter possibility, *S*s in both conditions were told that, although there was no shock for incorrect responses, the experimental session would continue until their performance satisfied a criterion of speed and accuracy.[1]

The stimuli were three squares having a projected size of .75, 1.12, and 1.50 in. square. These sizes were found to be equally spaced on the equal discriminability scale described by Garner and Hake (8). Between *S*s the large and small square were alternated as the training stimulus. The *S*s were run individually. Before starting the experiments, *S* was shown the three stimuli, one at a time, to insure that he could correctly identify each. Twenty training trials were then given in which the training stimulus was the only stimulus presented. Five test trials for the training stimulus and for each of the two generalization stimuli followed the training trials without interruption. An additional 42 training trials were interspersed among the test trials in such a manner that each test trial was separated from the other test trials by three training trials. Thus for an *S* who was trained on the large square, the experimental sequence would be: 20 successive presentations of the large square, a test presentation of the large, medium, or small square, three more presentations of the large square, and so on. Two counterbalanced sequences were employed with the stimuli in the test trials to minimize sequence effects in the generalization gradients.

Subjects. A shortened version of the MMPI, containing the items for the hysteria, psychasthenia, lie, and *K* scales as well as the items for Taylor's scale of manifest anxiety (17), was administered to 120 males enrolled in the introductory course in psychology. The *S*s used in the experiment were selected from among those who scored above the 80th decile (hysterics) and below the 20th decile (psychasthenics) on the distribution of scores obtained by subtracting each individual's score on the psychasthenia scale from his score on the hysteria scale. This subtractive procedure was used since it had been previously found to be effective (1, 6) as a way to partial out a general factor of "neuroticism" which contributes to high scores on both scales. A total of 18 hysterics, 12 in the avoidance condition and 6 in the nonavoidance, and 21 psychasthenics, divided 14 and 7 between the avoidance and nonavoidance conditions, constituted the experimental *S*s.

Apparatus. The *S* was seated at a table 2 ft. away from a 3 × 4-ft. vertical plywood partition. In the lower left side of the partition was a horizontal slot ½ in. wide and 11 in. long. In the right-hand side of the partition, at eye level, was an 8 × 8-in. ground glass window. The stimuli were presented to *S* by projecting them on the window. On a ready signal from *E*, *S* inserted a pencil into the slot on the left-hand edge where the pencil made contact with a microswitch. When the training stimulus appeared, *S* made a rapid horizontal movement of the pencil to the right, all the while keeping the pencil in the slot. The extent of the required movement was 23 cm. (approximately 9 in.), which was indicated by a marker on *S*'s side of the partition. The *S*'s response was recorded by the tracing of the pencil on recording paper mounted behind the partition. The stimuli were projected by a 35-mm. projector. Through the use of a condenser-time-delay circuit the duration of exposure of each stimulus was 3 sec.

[1] Although the above procedure is quite similar to that used by Rosenbaum (14), it differs from his in two important respects. Rosenbaum failed to provide a control against *S*s who may deliberately decide to respond to every stimulus, and he also did not inform *S*s that the shock circuit would not always be on when the training stimulus appeared. An attempt by the present experimenter to duplicate exactly Rosenbaum's procedure revealed that a number of *S*s deliberately responded to every stimulus since they "had nothing to lose" because errors were not punished. Some *S*s also discovered that the shocks were arbitrarily administered by *E* since, in accord with the arbitrary schedule, they would sometimes be shocked when they had made a very rapid response and not shocked when their response was unusually slow.

The latency of *S*'s response was measured to the nearest hundredth of a second by a Standard Electric Timer. The timer was activated simultaneously with the projector and was stopped when *S* broke contact with the microswitch at the left-hand edge of the slot. If *S* failed to respond, the timer automatically stopped when the projector circuit was interrupted by the timing-delay apparatus.

Electric shock was administered to the back of *S*'s left hand by hand electrodes. The source of shock was a faradic current from an inductorium supplied by a 6-v. dry cell. Strength of shock varied slightly between *Ss*, since the attempt was made to keep the strength of shock just below the level where *S* would refuse to continue with the experiment.

RESULTS

For each *S* the mean response amplitude on the test trials was computed for each of the three stimuli. The means of the individual *S* means are plotted in Fig. 1, by experimental condition, for the hysteric and psychasthenic groups. While there is essentially no difference in response amplitude between groups or conditions to the training stimulus (0), the amount of generalization to the two nontraining stimuli (stimulus 1 and 2) is consistently greater in the hysteric group under both experimental conditions. Also both the hysteric and psychasthenic groups show more generalization under the avoidance than under the nonavoidance condition. There is little difference, however, in the amount of generalization shown to stimulus 1 and 2.

Statistical analysis supports the graphical interpretation of the data. Since the variance of the response amplitudes was significantly less in the nonavoidance condition, it was not appropriate to combine the two experimental conditions in a common analysis of variance. Instead a modified three-way analysis (groups, the two generalization stimuli, and individuals) was performed separately for the avoidance and nonavoidance conditions. The results of these analyses are summarized in Table 1.

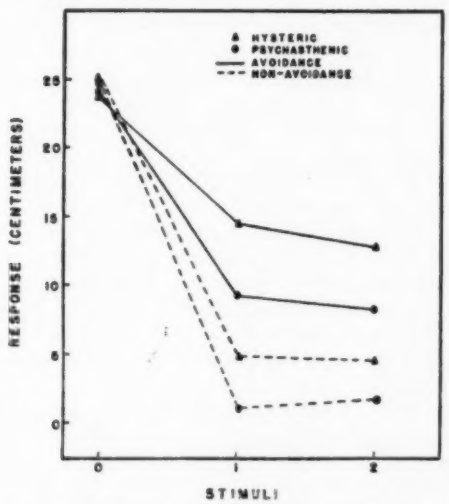


FIG. 1. MEAN RESPONSE AMPLITUDES TO THE TRAINING AND GENERALIZATION STIMULI

TABLE 1

SUMMARY OF THE ANALYSES OF VARIANCE FOR THE AVOIDANCE AND NONAVOIDANCE CONDITIONS

Source	Mean Square	*df*	s^2	*F*
Avoidance Condition				
Total	2024.35	51		
Groups	280.72	1	280.72	4.76**
Stimuli	26.90	1	26.90	2.15
Individuals	1415.32	24	58.97	
Groups × Stimuli interaction	.77	1	.77	.06
Pooled *Ss* × Stimuli interaction	300.64	24	12.53	
Nonavoidance Condition				
Total	350.17	25		
Groups	70.77	1	70.77	3.30*
Stimuli	.30	1	.30	.08
Individuals	235.72	11	21.43	
Groups × Stimuli interaction	.68	1	.68	.18
Pooled *Ss* × Stimuli interaction	42.70	11	3.88	

* Significant at the .05 level (one-tailed test).
** Significant at the .02 level (one-tailed test).

For both experimental conditions, differences between the hysteric and psychasthenic groups were significant beyond the .05 level for a one-tailed test in the predicted direction. In neither condition was there a significant difference between the generalization stimuli. Nor were the interactions significant. In Table 2 will be found the means of the individual *S* means by groups and experimental conditions along with their accompanying *t* tests.

Because of a significant difference in the variance of responses under the avoidance and nonavoidance conditions, it was not possible to compare the difference in over-all generalization between these two conditions by the standard *t* test. Instead the method suggested by Cochran and Cox (2) was used; this yielded a *t* value of 5.31, significant beyond the .001

TABLE 2

MEAN RESPONSE AMPLITUDES IN CENTIMETERS BY GROUPS AND EXPERIMENTAL CONDITIONS TO THE TRAINING AND GENERALIZATION STIMULI

Group	Avoidance Condition				Nonavoidance Condition			
	Stimuli				Stimuli			
	0	1	2	1 and 2 combined	0	1	2	1 and 2 combined
Hysteric	24.0	14.4	12.7	27.1	25.0	4.8	4.7	9.6
Psychasthenic	24.1	9.5	8.3	17.7	25.2	1.1	1.7	2.8
t	.01	2.17**	1.81*	2.18**	.42	2.28**	1.34	1.85*

* Significant beyond the .05 level (one-tailed test).
** Significant beyond the .02 level (one-tailed test).

TABLE 3

MEANS OF THE MEDIAN RESPONSE LATENCIES IN HUNDREDTHS OF A SECOND BY GROUPS AND EXPERIMENTAL CONDITIONS TO THE TRAINING AND GENERALIZATION STIMULI

Group	Avoidance			Nonavoidance		
	Stimuli			Stimuli		
	0	1	2	0	1	2
Hysteric	.26	.23	.23	.34	.26	.26
Psychasthenic	.30	.24	.26	.36	.35	.31

level for a two-tailed test. Thus it seems safe to conclude that the avoidance condition results in greater stimulus generalization.

The median latency for positive responses during the test trials was determined by *S*s for each stimulus. The means of these median latencies are given in Table 3 by group and experimental condition. Three trends can be noted in this table. First, the latencies to the generalization stimuli are shorter than to the training stimulus. Second, the response latencies for the hysterics are consistently shorter than for the psychasthenics, and third, the latencies for the avoidance condition are shorter than for the nonavoidance. While none of these trends is supported by statistical test, the shorter latencies for the responses to the generalization stimuli are similar to the findings of Gibson (9) and Rosenbaum (14). They are probably to be expected since the less time an *S* takes for discrimination, the greater the likelihood of a wrong response.

DISCUSSION

The above results are in good agreement with the prediction. The hysterics under both avoidance and nonavoidance conditions show a greater degree of stimulus generalization than the psychasthenics. The demonstration of a difference between these two groups, however, does not insure that our theoretical basis for expecting this difference was correct. While the prediction for the present study was arrived at from the assumption of a difference in the way hysterics and psychasthenics respond to the internal cues of anxiety, other explanations are readily available for the present results. Further study of such extreme groups is required before we can be confident as to the factors underlying the differences in performance.

Irrespective of what may be the basis for the obtained differences in stimulus generalization, the present findings are in keeping with a growing amount of experimental work showing the relation of the hysteria-psychasthenia variable to various aspects of performance under stress. In addition to the previously discussed relation of this personality variable to learning and memory performance, Davis (3, 4) has shown characteristic differences in performance on a psychomotor task between hysterics and psychasthenics when stress was induced by rapid pacing on the task. Results closely related to those of the present study are reported in an experiment by Deese, Lazarus, and Keenan (5). They found that *S*s high on the MMPI hysteria scale showed impairment in nonsense syllable learning under stress produced by electric shocks, whereas the performance of *S*s high on the psychasthenia scale was little affected.

The success of the hysteria-psychasthenia variable in predicting various aspects of performance under different stress conditions would appear to establish it as an important dimension of personality. As such there is a need for a clearer definition of the variable in personality and behavioral terms. In a previous article (6) the similarity of the hysteria-psychasthenia variable and Eysenck's factor of introversion-extroversion (7) was pointed out, but as yet there is insufficient evidence for positively identifying it with Eysenck's factor.

It is interesting to note that the composite hysteria-psychasthenia scale used in the present study is quite highly correlated with the Taylor scale of manifest anxiety ($r = -.72$, $N = 120$). The magnitude and direction of the correlation are such as to show that

Ss on the hysteria pole are low scorers on the anxiety scale and those on the psychasthenia pole are high scorers. Taylor (16) and Taylor and Spence (18) have shown that high- and low-anxiety groups selected on the scale of manifest anxiety differ in performance in conditioning and verbal maze learning. They have interpreted these results along Hullian lines with anxiety being considered as a generalized drive influencing performance through the building of habit strength.

The present experiment and that of Deese *et al.* (5) suggest that the Taylor and Spence interpretation of the anxiety scale is too simple an interpretation. If the drive interpretation of anxiety were correct, we would have expected the hysteric group in the present experiment to show the least stimulus generalization and the psychasthenic group the most. In fact, however, the opposite held true. This finding, together with the other evidence mentioned above, suggests that the Taylor scale of manifest anxiety is not so much a measure of anxiety as it is a measure of a personality variable relating to the way individuals respond to or express their anxiety.

SUMMARY

The amount of stimulus generalization under experimentally induced anxiety was determined in an hysteric and a psychasthenic group of *Ss*, selected on the basis of these respective scales in the MMPI. Three different size squares were used for the training and generalization stimuli, and the response was a simple horizontal arm movement. Two experimental conditions were used. Under the *avoidance* condition *Ss* were instructed that failure to respond rapidly would result in their receiving a strong electric shock. Under the *nonavoidance* condition *Ss* were told that they would receive electric shocks randomly, irrespective of their performance. In fact, however, both conditions had the same number of shocks administered in a predetermined order.

The hysteric group of *Ss* showed significantly more stimulus generalization under both experimental conditions. Also for both the hysteric and psychasthenic groups, the avoidance condition resulted in significantly more generalization. The results were considered in relation to a broad concept of stress, and it was suggested that the Taylor scale of manifest anxiety was not so much a measure of anxiety level as it was a measure of how *Ss* respond to or handle their anxiety.

REFERENCES

1. CARLSON, V. R. Individual differences in the recall of word-association test words. Unpublished doctor's dissertation, The Johns Hopkins Univer., 1953.
2. COCHRAN, W. G., & COX, G. M. *Experimental designs.* New York: Wiley, 1950.
3. DAVIS, R. Neurotic predisposition and the disorganization observed in experiments with the Cambridge Cockpit. *J. Neurol. Neurosurg. Psychiat.*, 1946, **9**, 119–124.
4. DAVIS, R. Increase in strength of a secondary drive as a cause of disorganization. *Quart. J. exp. Psychol.*, 1948, **1**, 22–28.
5. DEESE, J., LAZARUS, R. S., & KEENAN, J. Anxiety, anxiety reduction, and stress in learning. *J. exp. Psychol.*, 1953, **46**, 55–61.
6. ERIKSEN, C. W. Psychological defenses and "ego strength" in the recall of completed and incompleted tasks. *J. abnorm. soc. Psychol.*, 1954, **49**, 45–50.
7. EYSENCK, H. J. *Dimensions of personality.* London: Kegan Paul, 1947.
8. GARNER, W. R., & HAKE, H. W. The amount of information in absolute judgments. *Psychol. Rev.*, 1951, **58**, 446–459.
9. GIBSON, ELEANOR J. Sensory generalizations with voluntary reactions. *J. exp. Psychol.*, 1939, **24**, 237–253.
10. HILGARD, E. R., JONES, L. V., & KAPLAN, S. J. Conditioned discrimination as related to anxiety. *J. exp. Psychol.*, 1951, **42**, 94–100.
11. LAZARUS, R. S., DEESE, J., & OSLER, SONIA F. The effects of psychological stress upon performance. *Psychol Bull.*, 1952, **49**, 293–317.
12. MAHL, G. F. Anxiety, HCl secretion, and peptic ulcer etiology. *Psychosom. Med.*, 1950, **12**, 158–169.
13. MEYER, D. R. On the interaction of simultaneous responses. *Psychol. Bull.*, 1953, **50**, 204–220.
14. ROSENBAUM, G. Stimulus generalization as a function of level of experimentally induced anxiety. *J. exp. Psychol.*, 1953, **45**, 35–43.
15. ROSENZWEIG, S., & SARASON, S. An experimental study of the triadic hypothesis: reaction to frustration, ego-defense, and hypnotizability. I. Correlational approach. *Charact. & Pers.*, 1942, **11**, 1–19.
16. TAYLOR, JANET A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–93.
17. TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.
18. TAYLOR, JANET A., & SPENCE, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44**, 61–64.

Received July 23, 1953.

TACTUAL PERCEPTION OF FORM BY THE BRAIN-INJURED[1]

ALAN O. ROSS

Yale University[2]

Most of the tests used by the psychologist for the diagnosis of brain injury employ visual or auditory stimuli (8, 9, 17). The neurologist, on the other hand, frequently uses tactual stimuli for the examination of patients suspected of being brain-injured, particularly when the lesion is believed to be in the posterior parietal area, where injury is said to result in an inability to recognize forms tactually (astereognosis) (3, 6, 11). These neurological techniques, however, are rarely objective and never quantitative. It is the purpose of this study to determine whether an objective, quantitative tactual test of perception of raised outlines of forms can be used to discriminate the performances of brain-injured from those of normal human adults.

A test was designed on the principle that outlines of a form can be presented in varying degrees of completeness and that the degree of completeness can be expressed in terms of the number of tactual stimuli of which the outline is composed and the distance between these stimuli. Since brain-injured patients have been reported to exhibit impaired tactual perception, especially when the parietal lobe is involved, and since brain-injured patients are said to have difficulty in responding to two simultaneously presented stimuli (2, 5, 13, 17), two hypotheses concerning their performance on the tactual test may be advanced.

Hypothesis I. Patients with brain injury require a greater degree of completeness of a form outline than normals in order to recognize the form tactually.

Hypothesis II. Patients with brain lesions differentiate less well than normals between two forms presented simultaneously through the tactual modality.

[1] Part of a dissertation presented for the degree of Doctor of Philosophy at Yale University. The valuable advice and critical comments of Col. Charles S. Gersoni and the members of the doctoral committee, Drs. S. B. Sarason, H. E. Rosvold, and J. R. Wittenborn, are gratefully acknowledged.

[2] Now at Brooke Army Hospital, Fort Sam Houston, Texas.

METHOD

The Tests

The experiment consisted of two test series. In the tactual tests, the stimuli were thumbtack heads outlining letters, numbers, and well-known symbols on individual plywood boards, presented in one or two open-ended boxes which screened the boards from *S*'s view.[3]

Identification series. The main test made use of eleven forms which *S* had to identify. Six of the forms were 9 in. by 9 in., five were 4 in. by 4 in. in their largest dimension. Each form could be presented in four steps of decreasing order of difficulty. The first step of each form consisted of the outline of the form made up of the least number of tacks which could be used without destroying the essential characteristics of the form. On each subsequent step the outline consisted of more tacks until, with the fourth step, the form was completely outlined by adjoining tacks. The four steps thus differed in the number of tactual stimuli and the distance between these stimuli.

Each of the 11 forms of the main test was followed by a *tactual multiple-choice* recognition test bearing upon the form which it followed. One of these forms was a miniature wire replica of the form represented by the particular set, the other three were different forms which had some aspect, such as a distinct angle, in common with the form represented in the set.

A *visual multiple-choice* recognition test was presented once, at the end of the identification series. This test consisted of a card on which the 11 forms used in the main test were reproduced in line drawings. The 11 test forms were randomly scattered among 29 other forms from which they were to be chosen.

Discrimination series. This consisted of ten paired figures of which each pair was presented simultaneously. Using both hands, *S* had to say whether the two figures were the same or different. This series consisted of letters or numbers outlined with tacks, placed adjacent to one another. Five pairs contained identical forms, while the other five contained forms which were similar in some respect.

Procedure

Subjects were examined individually. Each *S* was told that the examination was in the nature of a survey,

[3] Details of the procedure as well as illustrations and descriptions of the stimulus boards, the boxes, and other material are contained in an appendix which has been deposited with the American Documentation Institute. Order Document No. 4164 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $2.00 for 35 mm. microfilm or $3.75 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

assurance being given that the results would not in any way bear upon his status as a patient or as a soldier.[4] He was then asked to indicate his dominant hand and to place this hand in one of the screening boxes. A brief tactual screening test was given in which *S* had to identify by touch three out of four items of everyday use.

Identification series, main test. For this test the instructions stressed the fact that each of the boards had a figure or a form on it, that these forms had a meaning which could be described, and that *S* was expected to tell what the figure was and to reproduce it by drawing. The *S* was assured that the form was always a single object which could be recognized if seen. Two sample boards were shown visually to acquaint *S* with the construction of the forms. A time limit of one minute was set for each board and *S* was encouraged to guess if necessary.

The first step of a form was placed in the box. As soon as *S* gave a response, or at the end of one minute, the board was withdrawn and *S* was asked to draw what he thought the form looked like. If *S* identified the form correctly,[5] either by naming or by drawing, or both, the *tactual multiple-choice* board was placed in the box and *S* was asked to examine the four figures and to point to the one which had the same shape as the main test form just completed. This was followed by the first step of the next form in the series.

An incorrect response or a failure to respond to a board on the main test was followed by the next step in the set of four boards until either a correct response was obtained or the time limit for the fourth step had expired, and *S* had thus failed to recognize the particular form on any of its four steps. In this case the next form in the series was presented, starting with the first step.

After completion of the last form of the main test, the *visual multiple-choice* recognition test was presented and *S* was instructed to point to all the forms he remembered having worked on. A time limit of half a minute was set for this test. This completed the identification series.

Discrimination series. For this series *S* was instructed to place one hand in each box and to state whether the two forms on the boards he was to feel were the same or different. Time limit for each pair was half a minute. If *S* delayed his response, he was encouraged after 23 seconds; after a response was obtained or the time limit had expired, the boards were replaced by the next pair in the series.

After completion of the discrimination series, *S* was given the Bender-Gestalt Test (1, 12) as part of another study and the CVS Individual Intelligence Scale (10). The CVS contains the Comprehension and Similarities subtests of the Wechsler-Bellevue scale (20) and a short vocabulary test, based on the Stanford-Binet word list (16, 18). The entire experimental procedure never lasted longer than 1 hour and 45 minutes.

Scoring

The main test of the identification series was scored by assigning a value of 4 to a correct response on the first step of a form; a value of 3 to a correct response on the second step; 2 for the third; and 1 for the fourth. Failure to give a correct response on all steps of a form earned the score of zero.[6] Since one form was used as an introduction and was not scored, the maximum total score for the main test of the identification series was 40, the minimum score zero. This score will be referred to as the *main score* in the following discussion.

On the tactual multiple-choice recognition test, a score of 1 was given for each correct choice, leading to a possible maximum score of 10. The visual multiple-choice recognition test was scored by giving 1 point for each design correctly recalled. The maximum score for this test was 11. Each of these subtests was scored and analyzed separately and no over-all total score was calculated for the series.

The discrimination series was scored by crediting 1 point for each correct discrimination, for a total possible score of 10.

Subjects

The *S*s were patients from the neurosurgical, neurological, and rehabilitation services of three Army hospitals. The test was administered to every patient whose major diagnosis was craniectomy, regardless of cause, and who was present during the week in which the hospital was visited. A total of 31 patients were examined. Seven of these had to be discarded before analysis of the data, leaving a sample of 24.[7] Discards were necessitated by the diffuse nature of some of the lesions which made localization impossible, by the patients' emotional state interfering with testing, or by the fact that diagnoses on which the initial selection had been based had implied surgical procedures which, upon closer study of the medical history, were revealed as not having taken place.

Detailed abstracts of the medical and operative histories of the patients were obtained and, wherever possible, the Army General Classification Test score (19) or the Armed Forces Qualification Test score was secured from personnel records. The site of the lesion was ascertained by close scrutiny of operative notes and from skull X rays, encephalograms, and surgeons' statements.

The normals were obtained from the service personnel stationed at an Army hospital. In order to find a match for each patient, controlling for intelligence, age, years of Army service, and years and geographic region of education, 42 normals were selected from the available population by reference to a personnel roster. These *S*s were screened for histories of trauma or disease which might have affected their brain structure.[8] Five cases had to be eliminated because some residual brain damage was suspected, leaving 37 normals who were tested.

From this pool of normals, a match for each of the 24 patients was drawn. The intelligence measure used for the matching was the CVS Individual Intelligence Scale (10). The data pertaining to the matching varia-

[4] The verbatim instructions are contained in the appendix (see footnote 3).

[5] For criteria see section on scoring and appendix (see footnote 3).

[6] Methods and criteria used in scoring are described in the appendix (see footnote 3).

[7] Detailed information on the 24 patients and on the discarded cases is contained in appendix Tables 1 & 2 (see footnote 3).

[8] See check list in the appendix (see footnote 3).

TABLE 1

MATCHING VARIABLES

(N = 24 pairs; variables matched for same pairs throughout)

Variable	Mean	σ	t	F	r
Intelligence					
Normals	29.58	6.20	.78	1.02	.99
Patients	29.42	6.25			
Age (years)					
Normals	24.5	4.5	.79	1.12	.75
Patients	23.9	4.8			
Education (years)					
Normals	10.9	1.9	.24	1.68	.74
Patients	11.0	2.5			
Service (years)					
Normals	3.6	2.7	.76	1.54	.74
Patients	3.3	2.2			

Note.—None of the differences is significant. All of the correlations are significant below the .01 level.

bles are summarized in Table 1. The mean difference t test and the F test for homogeneity of variance indicate that the two groups were probably drawn from the same population. The high product-moment correlations show that the individuals within the groups were closely matched.

Civilian occupations of normals and patients were similar in distribution,[9] so that a possible bias because of occupational experience with tactual tasks could be ruled out. All *S*s passed the tactual screening test, thus eliminating peripheral anesthesia as explanatory of the findings.

RESULTS

Identification series, main test. According to the principle on which the test was designed, the main score is an expression of the degree of completeness required by an S in order to recognize a tactually presented form outline. Should the patient group score significantly lower than the normal group, Hypothesis I, that patients with brain injury require a greater degree of completeness of a form outline than normals in order to recognize the form tactually, would be confirmed. The main-score mean for the patients is significantly below the mean for the normals with a p below .01 (Table 2), supporting Hypothesis I.[10]

Both verbal and graphic responses were credited as correct in order to avoid a handicap to the patients who might have had verbal difficulties of an aphasic nature. The number of *S*s in each group who drew the figures before naming them should indicate the extent of such a handicap. A chi-square analysis shows that there is no significant relationship between this manner of responding and membership in the patient group (chi square = .82, df = 2).

The lower main score of the patients might be the result of their slowness in responding. A comparison of the patients' mean response time per board with that of their respective controls indicates that there is no significant difference between the groups (Table 2). A further comparison, relevant to the question of slowness, deals with the number of times *S*s had boards withdrawn from them because of expiration of the time limit. A chi square of .003, df 2, indicates that there is no association between brain injury and slow performance. The time factor was apparently not responsible for the poorer performance of the patients.

Patients had significantly lower scores than normals when only the first five forms of the main test of the identification series are considered (Table 2). Since this includes no more than 20 minutes of testing time, the fatigue factor was apparently not responsible for the lower score of the patients on the test as a whole.

The correlation between the main scores and the CVS scores, which represent a post-injury intelligence measure for the patients, is .53 ($p < .01$) for the normals and .26 ($p > .10$) for the patients. The two correlations do not differ significantly ($z = 1.06$, $p = .29$).

TABLE 2

SUMMARY OF SCORES FOR IDENTIFICATION SERIES

Score	Mean		t
	Normals (N = 24)	Patients (N = 24)	
Main	25.1	20.1	3.74**
First 5 forms (main)	16.1	13.9	2.63**
Mean time per board	41.8	43.1	.75
Tactual multiple choice	8.6	7.9	1.50
Visual recognition			
Number right	5.4	4.0	3.60**
Number pointed to	7.7	5.5	5.39**
Number wrong	2.3	1.5	1.99
Accuracy ratio	172.9	173.0	.02

Note.—Tests of significance based on one tail of distribution for first three comparisons; on two tails for tactual multiple-choice and visual recognition data.

** Significant at the .01 level.

[9] See appendix Table 3 (see footnote 3).

[10] Since the hypotheses predict differences in one direction, one tail of the distribution of t was used in all comparisons deriving directly from the hypotheses. All other tests of significance are based on two tails of the distribution.

Tactual multiple-choice recognition test. The score on this test is a measure of immediate memory. The comparison reveals no significant difference between patients and normals (Table 2).

Visual multiple-choice recognition test. Presented at the end of the identification series, this test is a further measure of memory. Patients chose significantly fewer correct figures than normals (Table 2). The number of correct figures chosen is subject to the effect of several variables. The greater the total number of figures an *S* points to within the half minute available, the greater will be the probability of his including figures which are scored as correct. Normals pointed to significantly more figures than did the patients (Table 2). When the number of correct figures is weighted by the total number of figures pointed to, the resulting accuracy ratio fails to reveal a difference between patients and normals (Table 2). It appears that the greater number of correct figures chosen by the normals is a result of their pointing to more figures within the time available. It may be possible to correct for this in the future by requiring a limited number of choices. Normals also pointed to more wrong figures than patients, but this difference is not significant (Table 2).

Occasionally an *S* failed to recognize a form on all four steps of the identification series' main test. Patients tend to have more such failures than normals, as indicated by a chi square which approaches significance (chi square = 9.37, $df = 4$, $p > .05$). This raises the possibility that patients were not able to make as many correct choices as normals because they were familiar with fewer forms which they might recognize. Having failed to recognize the letter *R* tactually, for example, they were unable to recognize the *R* on the visual multiple-choice test. Accordingly, performance on the visual multiple-choice test was rescored considering only those forms which an *S* had recognized on either the main test or the tactual multiple-choice test of the identification series. The percentage scores so obtained indicate that normals performed significantly better than patients ($p < .01$ on Wilcoxon's paired replicates test [21]).

Discrimination series, main test. This test was designed as a test of Hypothesis II, regarding the differentiation between two simultaneously presented forms. Four patients were unable to perform on this test because they had only one functioning hand. The remaining 20 patients had a mean score of 6.8, compared to a mean score of 7.6 for the matched normals. The difference between the two means is significant at the .05 level with a t of 2.10, supporting Hypothesis II. Further analysis indicated that Hypothesis II is related to Hypothesis I. The discrimination scores correlate with the identification main scores with an r of .47 ($p < .05$) for the 24 normals and with an r of .48 ($p < .05$) for the 20 organics. It is possible to evaluate the difference between the groups on the discrimination series, holding tactual identification ability, as measured by the main score, constant. A ratio which measures performance on the discrimination series, independent of performance on the main test of the identification series, was obtained by dividing an *S*'s standard score for discrimination by his standard main score on identification. The mean difference for this ratio, patients versus normals, is not significant with a t of .25.

The parietal area is said to be critical for tactual recognition of form (6, 11). An attempt was, therefore, made to determine whether patients with parietal lesions required a more complete outline of forms than patients with lesions in other areas in order to recognize the form tactually. The patient group contained only three cases with lesions restricted to the parietal area, but a group of 10 patients could be formed when all other patients where the parietal area was involved in the injury, but not exclusively, were added to the purely parietal cases. Only four of these cases had an injury in the dominant hemisphere and used the dominant hand. The mean main score of this heterogeneous group of 10, who had some form of parietal involvement, is significantly lower than the mean main score of the remaining 14 patients where the parietal area was not involved in the injury ($t = 1.80$, $df = 22$, $p < .05$). The fact that the patients in the parietal group had injuries of various sizes implicating other areas to varying degrees makes it impossible to attach much weight to this result.

If tactual perception is considered as primarily localized in the parietal area, it might be

suspected that the presence of parietal cases in the sample of patients resulted in the low score of the patient group on the test on which the answer to Hypothesis I was based. In line with Hypothesis I, a new comparison of the main scores for the identification series was made, using only the 14 patients who had no parietal involvement. The mean for this group was 22.0, compared to a mean of 25.3 for their normal controls. The difference is significant at the .05 level ($t = 2.62$), indicating that the difference between patients and normals remains when all cases with parietal involvement are excluded so that the conclusion relevant to Hypothesis I is not changed.

Differences in test performance are frequently attributed to differences in the length of time since operation. The mean time since operation (not including cranioplasties) was 7.9 months, with a standard deviation of 3.3; the median was 8 months, the range from 2 to 15 months. To test for a possible dependence between time since operation and main score on identification series, a chi-square analysis was made, dichotomizing the patient group at the median of time since operation and at the median of identification main scores. A chi square of .001, $df = 1$, indicates that there is no significant tendency for long or short time since operation to vary with high or low main scores. When a Pearsonian correlation coefficient was computed for length of time since operation and main score, the nonsignificant r of .11 ($p = .70$) pointed to a similar conclusion.

Inasmuch as the patients took the Army classification test at the time of their induction into the service, the scores they obtained at that time represent a measure of their premorbid performance, while the CVS scores measure their performance after injury. The Army classification scores are based on one of two tests, the Army General Classification Test (AGCT) (19), or the Armed Forces Qualification Test (AFQT), which has superseded the AGCT in recent years. The two tests correlate on the order of .92, and the scores are comparable in terms of the performances measured. In the following, classification test scores will be referred to as "AGCT" scores, regardless of which of the two tests they were based on.

Alternate basis for matching. By matching the groups on the basis of the patients' post-injury intelligence scores it was shown that the lower tactual scores of the patients were independent of any general lowering of mental ability. In order to investigate differences in tactual scores between patients and normals when preinjury mental ability is held constant, a new matching was accomplished by pairing each patient with a normal who had the same "AGCT" score. Age and years of education were controlled for at the same time. Table 3 shows that the mean differences for the matching variables were not significant. When the principal comparisons relevant to Hypotheses I and II are repeated with preinjury intelligence held constant in this manner, the differences in identification main scores and discrimination scores remain significant at the .01 and .05 levels respectively.

TABLE 3

GROUPS MATCHED FOR PREMORBID INTELLIGENCE

Comparison	Mean		t
	Normals	Patients	
"AGCT" score	59.0	59.6	.62
Age	24.0	22.8	1.27
Education (years)	11.3	10.6	2.00
Identification (main)	25.8	20.4	3.43**
Discrimination	7.5	6.8	1.81*

Note.—$N = 20$ for each group, except for discrimination scores where $N = 16$ for each group. Tests of significance based on two tails of distribution for first four comparisons, on one tail for analysis of identification and discrimination scores.

* Significant at the .05 level.

** Significant at the .01 level.

DISCUSSION

A tactual test was designed on the principle that outlines of a form can be presented in varying degrees of completeness and that the degree of completeness may be expressed in terms of the number of tactual stimuli of which the outline is composed and the distance between these stimuli. Two hypotheses concerning the performance of the brain-injured on this test were advanced. The hypothesis that patients with brain injury require a greater degree of completeness of a form outline than normals in order to recognize the form tactually was confirmed. The hypothesis that patients with brain lesions differentiate less well than normals between two forms presented simultaneously through the tactual modality was also confirmed.

With the present form of the test, the scores of patients and normals have too great an overlap for the technique to be valuable for differential diagnosis of brain injury, but the study suggests that with greater refinement and extensive standardization, a test of this kind may be made clinically useful.

The results of the study are relevant to a number of theories, opinions, and practices current in the field of brain injury. One of these is the view that disturbed tactual recognition of form is associated with a lesion in the parietal area of the cortex (3, 6, 11). The data obtained in this experiment indicate that functions as complex as those required in the tactual test used can be interfered with by lesions in other areas as well. Inferior tactual ability can be found in patients who have undergone brain operations, grouped without regard to the cerebral area affected, although there is indication that patients with lesions involving the parietal area exhibit a greater decrement than patients with lesions in other areas.

It would appear that the technique employed in the present study is sensitive to disturbances of a general psychological function, under which tactual recognition of form can be subsumed. If the function of the intact cerebrum is viewed as one of integration or synthesis (4), the present findings might be interpreted as indicating a disturbance of the integrative function of the brain-injured subjects, manifesting itself in poor test performance.

The present findings are further relevant to the statements by several writers (2, 5, 13, 17) that patients with brain injuries experience difficulty when required to differentiate between two simultaneously presented stimuli. It was shown that normals perform better than patients when they had to differentiate tactually between two simultaneously presented forms. The patients' deficit on the discrimination task can, however, be reduced to the more basic deficit which operates in tactual recognition of form so that simultaneity, as such, does not seem to be of primary importance in reducing the performance level of brain-injured.

Impulsiveness is another characteristic occasionally ascribed to brain-injured (14, 15). The data for the visual multiple-choice recognition test show that normals are significantly more productive than patients in terms of the number of forms pointed to within the time limit, but the greater number of wrong forms chosen suggests a lack of precision on the part of the normals. While patients point to fewer forms, the forms they do point to are more often correct than the many forms pointed to by the normals, so that the relative accuracy of the two groups is the same. The evidence thus points against impulsiveness on the part of the patients. It has also been suggested that brain-injured patients restructure their environment on the level of their reduced capacities (7). Such a phenomenon may be involved in the findings on the visual multiple-choice task. Patients appear to have reduced the quantity of their performance, thus maintaining the quality.

It might be argued that the psychological reactions of the patients to the experience of brain surgery and hospitalization resulted in their lower tactual scores. Several factors refute such an explanation. The patients and the normals reacted to the testing situation in essentially the same manner. The instructions aimed at reducing test anxiety seemed to have had the desired effect on all but one patient, and he was eliminated. Besides this clinical impression, there is quantitative evidence which speaks against the presence of the so-called catastrophic reaction in the patients. Anxiety might slow down the performance of the patients so that they would exceed the time limit significantly more often than the normals or show a longer mean response time per board. This was not the case. The opposite effect, that patients were giving rash and indiscriminate responses in order to escape the situation, is disproved by the mean response time data and by the data on the visual multiple-choice test, where patients showed more accuracy and discrimination of choice than normals.

It might be argued alternatively that the tactual scores are an expression of intelligence and that they might have been predicted from a knowledge of intelligence scores based on more traditional tests. Since the differences between patients and normals predicted by the hypotheses prevailed both when preinjury and when postinjury intelligence was held constant, it would appear that the tactual scores are relatively independent of intelligence

scores. It is possible that for diagnostic purposes a knowledge of both tactual and intelligence scores would be useful.

SUMMARY

A tactual test of perception of raised outlines of forms was designed on the principle that such outlines can be presented in varying degrees of completeness and that the degree of completeness can be expressed in terms of the number of tactual stimuli of which the outline is composed and the distance between them. The power of this test to discriminate the performances of 24 brain-injured patients from a like number of normal adults was examined. Patients with brain injury were shown to require a greater degree of completeness of form outline than normals in order to recognize the form tactually. When two stimuli were presented simultaneously for tactual discrimination, patients performed worse than normals, but when tactual recognition of form was held constant, the simultaneity factor appeared to contribute little to the difference between patients and normals. It was suggested that the technique might be refined and standardized to make it useful for differential diagnosis of brain injury.

REFERENCES

1. BENDER, LAURETTA. Disturbances in visuomotor Gestalt function in organic brain disease associated with sensory aphasia. *Arch. Neurol. Psychiat.*, 1933, **30,** 514–537.
2. BENDER, M. B., SHAPIRO, M. S., & TEUBER, H. L. Allesthesia and disturbance of body scheme. *Arch. Neurol. Psychiat.*, 1949, **62,** 222–236.
3. BING, R., & HAYMAKER, E. *Compendium of regional diagnosis in lesions of the brain and spinal cord.* (11th Ed.) St. Louis: Mosby, 1940.
4. BRICKNER, R. M. An interpretation of frontal lobe function based upon the study of a case of partial bilateral frontal lobectomy. *Proc. Soc. Res. nerv. ment. Dis.*, 1934, **13,** 259–351.
5. CAMPORA, G. Astereognosis: its cause and mechanism. *Brain*, 1925, **48,** 65–71.
6. DEJONG, R. N. *The neurologic examination.* New York: Hoeber, 1950.
7. GOLDSTEIN, K. *The organism.* New York: American Book Co., 1939.
8. GOLDSTEIN, K. & SCHEERER, M. Abstract and concrete behavior: an experimental study with special tests. *Psychol. Monogr.*, 1941, **53,** No. 2 (Whole No. 239).
9. HALSTEAD, W. C. *Brain and intelligence, a quantitative study of the frontal lobes.* Chicago: Univer. of Chicago Press, 1947.
10. HUNT, W. A., & FRENCH, ELIZABETH G. The CVS Abbreviated Individual Intelligence Scale. *J. consult. Psychol.*, 1952, **16,** 181–186.
11. NIELSEN, J. M. *Agnosia, apraxia aphasia; their value in cerebral localization.* (2nd Ed.) New York: Hoeber, 1946.
12. PASCAL, G. R., & SUTTELL, BARBARA J. *The Bender-Gestalt Test; quantification and validity for adults.* New York: Grune & Stratton, 1951.
13. PORTEUS, S. D., & KEPNER, R. D. Mental changes after bilateral prefrontal lobotomy. *Genet. Psychol. Monogr.*, 1944, **29,** 3–115.
14. PORTEUS, S. D., & PETERS, H. N. Maze test validation and psychosurgery. *Genet. Psychol. Monogr.*, 1947, **36,** 3–86.
15. STRAUSS, A. A., & WERNER, H. Disorders of conceptual thinking in the brain-injured child. *J. nerv. ment. Dis.*, 1942, **96,** 153–172.
16. TERMAN, L. M., & MERRILL, MAUD A. *Measuring intelligence.* New York: Houghton Mifflin, 1937.
17. TEUBER, H. L. Neuropsychology. In R. E. Harris, *et al., Recent advances in diagnostic psychological testing; a critical summary.* Springfield, Ill.: Charles C Thomas, 1950. Pp. 30–52.
18. THORNDIKE, R. L. Two screening tests of verbal intelligence. *J. appl. Psychol.*, 1942, **26,** 128–135.
19. UNITED STATES WAR DEPARTMENT, ADJUTANT GENERAL'S OFFICE, CLASSIFICATION AND REPLACEMENT BRANCH, PERSONNEL RESEARCH SECTION. The Army General Classification Test. *Psychol. Bull.*, 1945, **42,** 760–768.
20. WECHSLER, D. *The measurement of adult intelligence.* (3rd Ed.) Baltimore: Williams & Wilkins, 1944.
21. WILCOXON, F. *Some rapid approximate statistical procedures.* Stamford, Conn.: American Cyanamid Co., 1949.

Received August 30, 1953.

THE CONSISTENCY OF INTERPERSONAL BEHAVIOR JUDGMENTS MADE ON THE BASIS OF SHORT-TERM INTERACTIONS IN THREE-MAN GROUPS[1]

ROBERT R. BLAKE, JANE SRYGLEY MOUTON, AND BENJAMIN FRUCHTER

The University of Texas

DIRECT interpersonal judgments made by strangers on the basis of short-term interactions may provide one basis for measuring individual personal and social characteristics of the kind that cannot be easily evaluated through objective tests. Is such behavior sufficiently consistent and can such judgments be made with sufficient reliability to justify their further consideration in connection with technical problems such as assembling individuals into effective work groups? Since previous studies (e.g., Barker [1], Bell and French [2], and Haythorn [3]) bear on this issue only indirectly, it seems desirable to seek a direct answer to these questions.

This paper reports an experiment designed to determine the extent to which one technique of direct assessment produces reliable judgments concerning the personal and social characteristics in individual behavior. The approach to estimating the reliability employed in this investigation is somewhat different from the usual method. In the more conventional reliability study conditions in the first administration are repeated in the second administration. In the present investigation the estimate of reliability obtained is dependent not only on the reliability of the judgments but also on the consistency of the behavior and similarity of the group structures in the two situations. Since these three sources of variation are confounded in the design, high consistency may be interpreted as indicating little variation in any of the three sources. In cases of low agreement, however, it cannot be determined which of the sources of variation account for it. The values obtained therefore can be considered estimates of the amount of consistency to be expected over situations where raters, tasks, and groups vary.

[1] This research was supported in part by the United States Air Force under Contract Number AF 18(600)-602, monitored by the 3305th Research and Development Group (Combat Crew Training Research Laboratory), Human Resources Research Center, Randolph Air Force Base, Texas. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United States Government.

EXPERIMENTAL DESIGN

The general plan called for direct assessment and observer evaluation of the behavior of each of 33 subjects who interacted as members of three-man groups in two sessions. Changes between the two group sessions included varying the task on which members worked, the composition of the groups, and the identity of the observers. These changes were introduced by the use of the following experimental plan.

The general plan shown in Table 1 called for eleven interaction groups for each of two tasks. Every group consisted of three members who were unacquainted with one another at the beginning of the two discussion periods, with the same individuals participating in different groups during both periods. That is to say, in the second discussion period the membership of the groups was changed so that the three who interacted with one another during the second period were also unacquainted with one another at the beginning of their interaction period. In addition the interaction was evaluated by trained observers whose schedules were arranged so that no observer watched the same person functioning in two different sessions.

The behavior of individuals composing each group was evaluated in two ways. At the end of both discussion periods the three members filled in a 12-item direct-assessment questionnaire describing their own behavior and that of the other two members. The observers also filled in the same questionnaire to describe the behavior they had observed in the members. The rankings received on these questionnaires constitute the basic data for assessing the consistency of behavior between and within the two discussion periods.

METHOD

Subjects

The *S*s were 33 college students. Two restrictions were imposed in selecting them. The first was that they

TABLE 1
EXPERIMENTAL PLAN USED TO INTRODUCE VARIATIONS IN TASK, GROUP COMPOSITION, AND IDENTITY OF OBSERVERS FOR TWO SESSIONS INVOLVING THE SAME 33 SUBJECTS

Session	Group No.			
	I	II	III	IV.....XI
1				
Members	A B C	D E F	G H I	J K L
Observer	1	2	3	4
2				
Members	A E I	B F J	C G K	D H L etc.
Observer	4	3	2	1

should not be acquainted with one another prior to the beginning of the study. The second was that they should be undergraduate students, a restriction introduced in order to insure homogeneity in the sample of *Ss* who were being asked to rate one another. Twenty-four of the *Ss* were men and nine were women. Groups were arranged so that men and women worked separately.

Observers

The observers (*Os*) were 11 members of a class in social psychology. They were pretrained for making these kinds of observations by participating as *Ss* in three previous discussion groups where they filled out questionnaires similar to the ones used in this study.[2]

Group Task and Procedure

Two tasks were used to produce group discussion. The groups were instructed to discuss the assigned topics, to come to agreement regarding a solution, and to write a brief report concerning the agreement reached. The first discussion was based on the question, "Should college grades be abolished?" The second task required the three *Ss* to construct one meaningful sentence which contained the following three words, *to*, *too*, and *two*. The second task was not so satisfactory as the first for producing interaction. After 15 minutes of interaction on the first task the groups were broken up and each individual was given a questionnaire. It included 12 items in terms of which each participant could rank himself and the two other members. Items were selected to permit assessment of personal and social characteristics that appeared in the small group discussions. In each case *S* was required to rank himself as one of the three members. The items used are shown in Table 2.

On completion of this phase, new three-man groups were created so that each individual interacted with two individuals different from those who had been members of his first group. After a 15-minute discussion period on the second task these groups were broken up and the same questionnaire was administered to them. After both sessions, *Os* filled in the same questionnaire to judge the personal and social characteristics of the three group members. Each *O* ranked three different people on each of the two occasions, making their rankings without knowledge of the rankings made by the group members themselves or by other *Os*.

Control of the previously mentioned factors precludes the possibility of obtaining high consistency between rankings due to the introduction of factors such as previous acquaintance, recall on the second occasion of rankings made of other members on the preceding occasion, or continuity in performance based on discussion of the same topic. Significant agreement between rankings can be interpreted as due to consistency in the judgments of the behavior of *Ss* as evaluated by the technique employed. Absence of significant agreement cannot be interpreted as an inconsistency of judgments since the lack of agreement may be due to differences in the tasks or structure of the groups.

Statistical Treatment of Data

The distributions of rankings were trichotomized and chi square was used as the statistical test in making the various item comparisons, such as between administrations, between *Ss*, between *Os*, and so on, and the contingency coefficient (*C*) was computed as the index of the degree of relationship. A limitation of the scoring method is that the rankings within any one group are not independent of one another since ties were not permitted. The contingency coefficient was preferred to the Pearson product-moment coefficient since it seemed improbable that the distribution characteristics needed for applying the latter method, i.e., continuous normal distributions, were obtained in the data. They do meet the conditions for applying the chi-square test, however, with the exception of low expected frequencies in some cells.

RESULTS

The Consistency of Judgments between Sessions

The design of the study permits determination of the consistency of judgments between discussion periods by comparing rankings received on the questionnaire during the first period with those received during the second. It will be recalled that the people *Ss* interacted with on the second occasion were different from those on the first, and that no *S* was acquainted with the other two members at the beginning of the two interaction periods. Furthermore, the topic of discussion was different.

The consistency of judgments between sessions can be looked at from different points of view including the individual's evaluation of his own behavior (self-self), the evaluation of his behavior by two other group members with whom he was participating (member-member), and the *Os'* evaluations of his behavior (observer-observer). In addition the combined

[2] The class members who conducted the sessions and served as *Os* were Richard Austin, Dorothy Babin, Robert Bell, John Bowen, Jack Brackeen, Jack Hain, Julian Hill, Bette Perry, Glenn Rosson, Alice Smith, and Jane Woodbridge.

TABLE 2

ITEMS USED FOR THE DIRECT ASSESSMENT OF INDIVIDUAL BEHAVIOR

Instructions: This questionaire is to be filled out as you saw the members of your group interacting during this period. For each item you are to rank the members of the group in order of preference. Use the number of each person to identify the members. Include yourself in the ranking.

Item	Abbreviation used in text
1. Rank the group members in terms of who contributed most to the group decision.	Contribution to group decision
2. Rank the group members in terms of who did most of the clerical work.	Clerical work
3. Rank the group members in terms of who most often functioned as the leader.	Leadership
4. Rank the group members in terms of who did most to keep the group "on the ball."	On ball
5. Rank the group members in terms of who was most agreeable and willing to go along with what the others wanted.	Agreeable and willing
6. Rank the group members in terms of who was the most dominant in the group situation.	Dominant
7. Rank the group members in terms of who most wanted to do a good job.	Good job
8. Rank the group members in terms of who was the most cooperative.	Cooperative
9. Rank the group members in terms of who was most satisfied with the group decisions.	Satisfied with group decision
10. Rank the group members in terms of who was most frustrated by the group situation.	Frustrated
11. Rank the group members in terms of who showed the most interest.	Interest
12. Rank the group members in terms of who yielded most to others.	Yielded

TABLE 3

CONSISTENCY OF JUDGMENTS BETWEEN SESSIONS

Item	Self-Self			Member-Member			Observer-Observer			Group-Group		
	x^2	p	C	x^2	p	C	x^2	p	C	x^2	p	C
1. Contribution to group decision	3.60	—	.31	9.57	.05	.48	7.09	.20	.32	7.07	.20	.42
2. Clerical work	10.01	.05	.48	4.61	—	.35	7.64	.20	.33	12.31	.02	.52
3. Leadership	20.92	.01	.62	9.33	.10	.47	10.90	.05	.50	9.48	.10	.47
4. On ball	10.39	.05	.49	6.37	.20	.40	4.36	—	.34	18.38	.01	.60
5. Agreeable and willing	3.05	—	.29	5.57	—	.38	5.99	.20	.39	6.12	.20	.40
6. Dominant	13.18	.02	.53	3.76	—	.32	10.36	.05	.49	19.41	.01	.61
7. Good job	7.14	.20	.32	4.51	—	.35	2.18	—	.25	4.42	—	.34
8. Cooperative	8.16	.10	.45	1.80	—	.23	11.99	.02	.52	3.13	—	.29
9. Satisfied with group decision	10.22	.05	.49	2.85	—	.28	2.73	—	.28	2.39	—	.26
10. Frustrated	5.25	—	.37	6.35	.20	.40	3.81	—	.32	6.94	.20	.42
11. Interest	11.45	.05	.51	9.71	.05	.48	3.81	—	.32	2.68	—	.27
12. Yielded	5.04	—	.36	5.17	—	.37	2.18	—	.25	2.83	—	.28

ratings of the three group members (the *S*'s self-judgments and the judgments by the two other members) were used as a basis for evaluating the individual's behavior on a group basis (group-group). The statistical summaries for these comparisons are shown in Table 3.

An examination of Table 3 shows that there is some consistency in judgments as assessed by certain items on the questionnaire. By way of summarizing the findings reported in Table 3, it seems clear that reliable judgments about characteristics of individual behavior can be made even on the basis of extremely short periods of interaction. Items like *leadership* (No. 3), *interest* (No. 11), and *dominance* (No. 6) were reliably judged in a number of the comparisons made. This finding supports the view that direct judgments about personal and social characteristics can be used as measures of certain aspects of behavior under conditions of short-term interaction when conditions in the first administration are different from those in the second administra-

TABLE 4

CONSISTENCY OF JUDGMENTS WITHIN EACH OF THE TWO SESSIONS

Item	Observer-Member (1)			Observer-Member (2)			Observer-Group (1)			Observer-Group (2)		
	χ^2	p	C	χ^2	p	C	χ^2	p	C	χ^2	p	C
Contribution to group decision	27.27	.01	.67	11.48	.05	.51	29.03	.01	.68	15.46	.01	.57
Clerical work	18.54	.01	.60	14.68	.01	.56	33.27	.01	.71	24.44	.01	.65
Leadership	26.72	.01	.67	14.94	.01	.56	18.85	.01	.60	20.93	.01	.62
On ball	18.88	.01	.60	5.48	—	.38	16.92	.01	.51	5.96	—	.39
Agreeable and willing	2.00	—	.24	5.50	—	.39	10.12	.05	.48	9.61	.05	.48
Dominant	16.62	.01	.58	5.13	—	.37	12.05	.02	.52	14.64	.01	.55
Good job	.55	—	.13	8.72	.10	.46	1.79	—	.23	4.05	—	.33
Cooperative	1.73	—	.22	2.48	—	.26	8.28	.10	.45	6.47	.20	.40
Satisfied with group decision	4.20	—	.34	2.68	—	.27	2.70	—	.28	1.76	—	.23
Frustrated	2.35	—	.26	5.72	—	.38	5.17	—	.37	2.30	—	.26
Interest	1.33	—	.20	5.33	—	.37	6.86	.20	.42	7.59	.20	.43
Yielded	8.93	.10	.46	14.40	.01	.55	2.68	—	.27	8.04	.10	.44
	Self-Observer (1)			Self-Observer (2)			Self-Member (1)			Self-Member (2)		
Contribution to group decision	24.28	.01	.65	16.68	.01	.58	10.67	.05	.49	9.99	.05	.48
Clerical work	34.10	.01	.71	24.60	.01	.65	21.59	.01	.63	14.14	.01	.55
Leadership	15.61	.01	.57	24.37	.01	.65	11.90	.02	.52	12.88	.02	.53
On ball	23.75	.01	.65	1.60	—	.22	18.46	.01	.60	2.51	—	.27
Agreeable and willing	9.65	.05	.48	3.80	—	.32	7.02	.20	.42	15.03	.01	.56
Dominant	7.24	.20	.42	13.30	.01	.54	6.44	.20	.40	7.20	.20	.42
Good job	4.33	—	.34	4.67	—	.35	3.19	—	.30	3.15	—	.30
Cooperative	6.75	.20	.41	7.12	.20	.42	4.50	—	.35	7.75	.20	.44
Satisfied with group decision	.40	—	.11	.42	—	.11	2.39	—	.26	.97	—	.17
Frustrated	8.61	.10	.46	3.12	—	.29	4.37	—	.34	2.96	—	.29
Interest	5.41	—	.38	8.29	.10	.45	1.58	—	.21	5.84	—	.39
Yielded	5.99	.20	.39	4.90	—	.36	3.26	—	.30	6.94	.20	.42

tion. This generalization seems to be true regardless of whether judgments are made by *Os*, by group members, or by *S*s themselves.

On the other hand judgments concerning other personal and social characteristics such as *yielded* (No. 12), *desired to do a good job* (No. 7), and *satisfied with the group decision* (No. 9) were not made reliably on a between-session basis. While these data demonstrate the feasibility of using direct-assessment techniques to judge certain personal and social characteristics, further experimental work is needed to determine what kinds of characteristics are most amenable to reliable observation.

The Consistency of Judgments within Sessions

The second approach concerns the degree of agreement between different judges looking at the same sample of behavior. One question here is concerned with the relationship between *Os*' rankings and judgments of each individual by the other two group members. Another relates to rankings given by the group as a whole to each member as contrasted with the judgments given by the *Os*. The difference between these two comparisons is that in the latter case the *S*'s self-judgment is included. The same types of questions can be asked by comparing the self-judgments with the *Os*' judgments and by comparing the self-judgments with the judgments given by the other two group members. The statistical summaries for each of these comparisons are presented in Table 4.

An examination of Table 4 indicates considerable agreement in the assessment of behavior on a within-session basis. This agreement appears when comparisons are made

TABLE 5

SUMMARY OF LEVELS OF SIGNIFICANCE OF ITEMS BETWEEN AND WITHIN SESSIONS

Item	Between Sessions				Within Sessions							
	Observer-Observer	Member-Member	Self-Self	Group-Group	Observer-Member		Observer-Self		Member-Self		Observer-Group	
					Session 1	Session 2	Session 1	Session 2	Session 1	Session 2	Session 1	Session 2
Contribution to group decision	II	I	III	II	I	I	I	I	I	I	I	I
Clerical work	II	III	I	I	I	I	I	I	I	I	I	I
Leadership	I	II	I	II	I	I	I	I	I	I	I	I
On ball	III	II	I	I	I	III	I	III	I	III	I	III
Agreeable and willing	II	III	III	II	III	III	I	III	II	I	I	I
Dominant	I	III	I	I	I	III	II	I	II	II	I	I
Good job	III	III	II	III	III	II	III	III	III	III	III	III
Cooperative	I	III	II	III	III	III	II	II	III	II	II	II
Satisfied with group decision	III	III	I	III	III	III	III	III	III	III	III	III
Frustrated	III	II	III	II	III	III	II	III	III	III	III	III
Interest	III	I	I	III	III	III	III	II	III	III	II	II
Yielded	III	III	III	III	II	I	II	III	III	II	III	II

between judgments given by *O*s and members, between members' judgments and self-judgments, between *O*s' judgments and self-judgments, and between judgments by *O*s and the group as a whole.

The tabulation also indicates that considerable agreement obtains between the judgments made by independent *O*s for a single interaction situation. Items that produce consistently reliable judgments are those involving *contribution to the group decision* (No. 1), *clerical work* (No. 2), and *leadership* (No. 3). On the other hand *satisfied with group decision* (No. 9) and *frustrated* (No. 10) were the least reliably judged of any in Table 4. The remaining items were judged consistently in some comparisons but not in others.

Of particular interest is the extent of agreement between the self-judgments on the one hand and those by the *O*s and by other group members on the other. Self-judgments are usually not utilized in analytic situations of this type, but apparently they agree well with the ratings made by others in regard to certain aspects of behavior such as *contribution to the group decision* (No. 1), *clerical work* (No. 2), *leadership* (No. 3). This statement seems to hold for both task situations used in the present study.

In order to provide a comparative analysis of the value of items from both a between-session and a within-session point of view a summary of the significance levels for the different comparisons on an item-by-item basis is presented in Table 5. The items are grouped into categories on the basis of their levels of significance. Category I contains all items significant beyond the 5 per cent level. Category II includes items significant between the 5 per cent and the 20 per cent levels. Category III contains the items that fail to achieve the 20 per cent level of significance. The items in the first category reach an acceptable standard for rejecting the null hypothesis. The second category contains items which, while not demonstrated to be significant within the limited sampling of the present study, merit further consideration because of their possible value for scaling behavior of the type under analysis. Items for which there is no demonstrated significance from one interaction period to another are included in the third category.

An inspection of this table shows, for example, that the item dealing with *contribution to group decision* (No. 1) produced reliable judgments on a within-session basis and that the judgments for this item are also reliably assessed on a member-member basis from a between-session standpoint. The same kind of statement can be made for the *leadership* (No. 3) and *dominant* (No. 6) items. On the other hand, items such as *good job* (No. 7) and *satisfied with group decision* (No. 9) do not

seem to be judged consistently on a within-session basis, nor do they produce consistent judgments from a between-session standpoint. A more detailed examination of this table will identify the items likely to be useful in the evaluation of behavior samples on a short-term interaction basis for the types of situations used in this study.

SUMMARY AND CONCLUSIONS

This paper dealt with the reliability of social judgments based on direct assessment of *S*s engaged in short-term interaction. In order to assess the reliability of these kinds of judgments an experimental design was devised in which the conditions in the two administrations were different.

The general plan called for the direct assessment and observer evaluation of the social and personal characteristics of the same 33 *S*s in two completely independent interaction situations. Changes between the two group situations included the task on which the members worked, the composition of the groups, and the identity of the *O*s. On completion of each task, group members were asked to fill in a 12-item, direct-assessment questionnaire on which they judged their own behavior and that of the other two group members on a ranking basis. The *O*s judged the behavior of the group members using the same questionnaire. The results confirm the proposition that reliable judgments of short-term interaction can be made even when the two situations are different. Items that permit consistent assessment from both the between- and within-sessions points of view include *leadership* (No. 3), *contribution to group decision* (No. 1), and *dominance* (No. 6). Also of interest is the finding that *S*s' judgments of their own behavior, a type of data frequently disregarded, are related significantly to the judgments made by *O*s and other group members and therefore merit further investigation for possible utility. These results support the view that direct-assessment data can serve as one basis for evaluating personal and social characteristics that cannot be satisfactorily measured by other techniques.

REFERENCES

1. BARKER, R. G. The social interrelations of strangers and acquaintances. *Sociometry*, 1942, **5**, 169–179.
2. BELL, G. B., & FRENCH, R. L. Consistency of individual leadership position in small groups of varying membership. *J. abnorm. soc. Psychol.*, 1950, **45**, 764–767.
3. HAYTHORN, W. The influence of individual members on the characteristics of small groups. *J. abnorm. soc. Psychol.*, 1953, **48**, 275–284.

Received March 17, 1954. Prior publication.

CASE REPORT

A BLIND ANALYSIS OF A CASE OF MULTIPLE PERSONALITY USING THE SEMANTIC DIFFERENTIAL

CHARLES E. OSGOOD AND ZELLA LURIA

University of Illinois

FOR MORE than a year the writers have been collecting semantic data from patients undergoing psychotherapy. This has been feasible through the cooperation of psychotherapists in various parts of the country. When the manuscript of "A Case of Multiple Personality" (3) arrived, the editor of this journal, without our knowledge, suggested to Thigpen and Cleckley that it would be interesting to have semantic data from each of the "personalities" in their patient. Thigpen and Cleckley accepted the suggestion and administered a form of the differential we have been using for this purpose to their patient in each of her three personalities. The editor also suggested that we might see how much these semantic data would allow us to infer about the patient without our having any knowledge of the case history, protocol, or prognosis.[1] This appeared to be a useful and rather intriguing way to estimate the validity and sensitivity of this instrument. If we could infer descriptions of the three personalities which correspond with clinical observations, and if we could make reasonably accurate interpretations and predictions, we would be encouraged to continue our efforts to improve the semantic differential as a clinical tool.

THE SEMANTIC DIFFERENTIAL

The semantic differential is a combination of association and scaling procedures designed to give an objective measure of the connotative meaning of concepts. The underlying logic (1) can be summarized as follows: (*a*) The process of description or judgment can be conceived as the allocation of a concept to a set of experiential continua defined by pairs of polar terms. Thus the connotative meaning of a linguistically complex assertion, such as "My father has always been a rather submissive person," can be at least partially represented as

MY FATHER active__:__:__:__:X:__:__passive
MY FATHER soft__:X:__:__:__:__:__hard

The greater the strength of association, e.g., ". . . extremely submissive, a regular doormat," the more polarized, toward 1 or 7, the allocation (2). Since many scales of judgment are highly intercorrelated (e.g., *good-bad*, *fair-unfair*, *honest-dishonest*, *kind-cruel*, and so forth all reflect mainly the single "evaluative" factor in judgments), a limited number of such continua can be used to define a semantic space within which the connotative meaning of any concept can be specified. This clearly indicates some variant of factor analysis as the basic methodology in developing such an instrument. Two such analyses have been completed, both providing evidence for three general factors, "evaluation," "potency," and "activity," and some unknown number of specific factors that are probably denotative in nature.

The form of semantic differential we have been using in studying psychotherapy is based on this factor analytic work. In the 10 scales used, it gives approximately equal weight to the first three factors isolated. These scales and their factor loadings are given in Table 1. The 15 concepts used in this form of the differential were selected after consultation with clinicians and pretesting for their differentiating power. Ideally, they should sample the major persons and problems involved in therapy-in-general; we are not entirely satisfied with the present set, however, and more work should be done here. The concepts used are also shown in Table 1. In the test form itself, concepts are rotated against scales in such a way that each con-

[1] We wish to thank Dr. J. McV. Hunt, editor of this journal, for utilizing the opportunity presented by the work of Thigpen and Cleckley to arrange for the data in this study, for suggesting the blind analysis, and for his general encouragement. We are also very grateful to Drs. Thigpen and Cleckley for giving the semantic differential to their patient and for their interest in the instrument.

cept appears once with each scale, but with a maximum interval between successive appearances of both. The subject is instructed to do his checking rapidly, without struggling over particular items, to give his "immediate impressions." A 150-item form such as this usually takes less than 10 minutes to complete.

Reordering the raw data for a single subject on a single testing yields a matrix of N columns (here, 15 concepts) and i rows (here, 10 scales). The *meaning* of a particular concept to the subject, as defined by the operations of measurement here, is the profile of numbers in its column (or, more efficiently, the position in the n-dimensional space defined by the projection of these numbers onto the factors). *Difference in meaning* for two concepts is defined by the distance between their positions in this space, as computed by the generalized distance formula, $D = \sqrt{\Sigma d^2}$, in which d is the difference in allocation of the two concepts on a single scale (2). The more similar any two concepts are in connotative meaning, the smaller will be the value of D. *Change in meaning* (of the same concept at different times during therapy, or in different "personalities") can be defined by the same operation, except that d here refers to the differences in allocation of the same concept on the same scale at different testings. The mathematical properties of this formula also allow us to represent the *semantic structure* of an individual in a concise form; computation of the distance, D, of every concept from every other concept yields an N/N matrix (here, 15/15) of distances which have the property of plotting within a space having dimensionality equal to the number of factors. To the extent that the individual subject being studied uses the same three factors isolated in our general factor work, his data will plot accurately in three dimensions.

TABLE 1

CONCEPTS AND SCALES USED IN THIS ANALYSIS

Concepts		
LOVE	MENTAL SICKNESS	SELF-CONTROL
CHILD	MY MOTHER	HATRED
MY DOCTOR	PEACE OF MIND	MY FATHER
ME	FRAUD	CONFUSION
MY JOB	MY SPOUSE	SEX

Scales and Their Factor Loadings			
Scales	Evaluation	Activity	Potency
valuable-worthless	.79	.13	.04
clean-dirty	.82	.03	−.05
tasty-distasteful	.77	−.11	.05
fast-slow	.01	.70	.00
active-passive	.14	.59	.04
hot-cold	−.04	.46	−.06
large-small	.06	.34	.62
strong-weak	.19	.20	.62
deep-shallow	.27	.14	.46
tense-relaxed	−.55	.37	−.12

THE SEMANTIC DATA

At this point we should state exactly what information we have about this case. We know that we are dealing with a case of triple personality, and these have been labeled for us (presumably by the therapists who collected the semantic data) "Eve White," "Eve Black," and "Jane." We suppose that the "White" and "Black" have some connotative significance—certainly, as will be seen, the quantitative semantic data distinguish sharply between them. We also know, of course, that the patient is a woman, presumably participating in some kind of therapy; we do not know the stage of therapy or whether or not she is hospitalized. We considered it also fair to ask (from J. McV. Hunt) about the following items of sociological status, because they contribute to the meaningful interpretation of certain concepts: Concept CHILD—does this woman have a child? Yes, she does. Concept SPOUSE—is this woman married? Yes, she is. Concepts FATHER and MOTHER—are her parents alive? The mother is, but Hunt doesn't know about the father. Concept MY JOB—has this woman had a job outside of homekeeping? Yes, she has. This is the sum total of our external information about the case.

The semantic differential was given to this woman twice while "in" each of her three personalities; a period of about 2 months intervened between the two testings. The raw semantic data for each of the three personalities are given in Tables 2–4. The roman numerals I and II refer to first and second

testings respectively.[2] We take the space to give these raw data in full so that anyone who is interested may study them and test any particular "hypotheses" about the case he wishes. Low values in these tables indicate judgments toward the polar term on the left and high values judgments toward the polar term on the right. The data in these tables show the semantic profiles, or meanings, of each concept at each testing and while "in" each personality. Since the form given at each testing was actually a double form (each item repeated once), we were able to estimate the reliability of these data. The immediate test-retest reliability coefficients for each of the testings are as follows: Eve White I, .82; Eve White II, .90; Eve Black I, .65; Eve Black II, .89; Jane I, .89; Jane II, .94. These coefficients indicate (*a*) a generally satisfactory level of reliability, (*b*) a consistent trend in all three personalities toward greater stability through time, and (*c*) that Jane is the most consistent or stable personality over short intervals of time and Eve Black is the least.

To obtain measures of semantic similarity and structure, we computed the matrices of *D* for each concept with every other concept, for each personality and testing. With an ordinary desk calculator and a table of square roots, these operations are very simple and rapid. In order to conserve space, the six matrices of *D* are not given here. These "distances" are based on application of the formula given earlier across all 10 scales. For convenience in plotting the models which appear as Fig. 1–6, the data for scales contributing to each of the three factors were averaged and new *D*'s computed. This, in effect, forces those data into three dimensions and, hence, into solid models that have no error. The very slight amount of distortion, or loss of information, resulting from this averaging process and restriction to three dimensions can be seen from the following correlations between original (10 scale) and "factor" *D* matrices (3 average scales): Eve White I, .91; Eve White II, .93; Eve Black I, .96; Eve Black II, .98; Jane I, .86; Jane II, .92. In other words, nearly all of the variance in this woman's judgments can be accounted for in terms of three factors. Figures 1–6, then, provide quite accurate representations of the ways various concepts are related in each of the personalities; the smaller the distance between any two concepts the more similar in connotative meaning they are.

[2] These values are actually averages of two checkmarks, a double form of the semantic differential having been used for reliability purposes. In this double form, each item appears twice.

THE THREE PERSONALITIES AND THEIR CHANGES THROUGH TIME

The general assumption we are following is that "mental illness" is essentially a disordering of meanings or ways of perceiving from those characteristic of people judged "normal" in our society, and that the process of psychotherapy from the patient's point of view is essentially a reordering and changing of these meanings. Within the limitations of our type of measurement and our sampling of concepts, the locations and relations among concepts shown in Fig. 1–6 can be thought of as pictures of how this woman perceives herself, the significant people about her, and certain modes of action—when functioning "in" her several personalities.[3] We assume that this woman is receiving some kind of treatment through the period covered by our two samplings, I and II, and therefore look particularly for the types of changes in meaning that are taking place in the three personalities, as well as at the general nature of their organization. For purposes of ready comparison, all of the models are oriented in respect to the concept MY DOCTOR, which stays practically constant in meaning (*good*, *strong*, and *quite active*) through both time and personalities; spatially, in the figures, *good* is up and *bad* down, *active* to the left and *passive* to the right, and *strong* is away from the viewer while *weak* is near to or toward the viewer; the solid ball represents the origin of the space, e.g., a hypothetical "meaningless" concept that would result from checking all 4's.

Eve White

Semantic structures for Eve White I and II are shown in Fig. 1 and 2. The most general characterization would be that *Eve White perceives "the world" in an essentially normal fashion, is well socialized, but has an unsatis-*

[3] The authors wish to thank Professor Jozef Cohen for his help in preparing these figures.

TABLE 2

RAW DATA FOR EVE WHITE

	LOVE		CHILD		MY DOCTOR		ME		MY JOB		MENTAL SICKNESS		MY MOTHER		PEACE OF MIND		FRAUD		MY SPOUSE		SELF-CONTROL		HATRED		MY FATHER		CON-FUSION		SEX		
	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II	
cold	5.0	5.5	4.0	6.0	6.0	5.0	4.0	3.5	5.0	4.0	4.0	5.0	1.5	6.0	5.5	4.0	1.0	1.0	4.0	2.5	4.0	5.0	1.0	1.0	6.0	5.0	1.0	4.0	3.5	2.0	hot
valuable	1.0	1.0	1.0	1.0	1.0	1.0	4.0	6.0	1.0	1.0	1.0	4.0	1.0	1.0	1.0	1.0	7.0	7.0	3.0	3.5	1.0	1.0	7.0	7.0	1.0	1.0	3.0	5.5	2.5	4.5	worthless
tense	4.0	2.5	4.0	7.0	7.0	7.0	1.0	1.0	2.0	2.0	1.0	1.0	1.5	3.5	7.0	6.5	1.0	1.0	4.0	5.5	5.0	3.0	1.0	1.5	1.5	6.0	1.0	1.0	4.5	2.5	relaxed
small	7.0	6.0	1.0	7.0	7.0	7.0	7.0	1.0	2.0	5.5	6.5	7.0	6.5	5.5	7.0	7.0	7.0	1.0	1.0	5.0	5.5	6.5	6.5	1.0	1.0	7.0	7.0	7.0	5.0	6.0	2.5	large
fast	7.0	4.5	2.5	2.0	3.0	2.0	5.5	6.0	5.5	4.0	7.0	6.0	1.0	2.0	7.0	6.0	1.0	2.0	4.0	4.0	6.0	6.0	7.0	3.0	4.0	2.0	1.0	2.0	5.0	4.0	slow
dirty	7.0	7.0	7.0	7.0	7.0	7.0	5.5	6.0	7.0	7.0	5.5	6.5	7.0	7.0	7.0	7.0	1.0	1.0	5.0	4.5	5.0	6.5	1.0	1.0	7.0	7.0	1.5	6.0	5.5	3.0	clean
weak	7.0	5.0	1.0	1.0	7.0	7.0	1.0	1.0	5.5	6.0	1.0	3.0	7.0	6.5	7.0	7.0	1.0	1.0	4.0	3.5	7.0	6.0	1.0	1.0	7.0	7.0	1.0	1.5	4.0	2.5	strong
tasty	1.0	2.0	1.0	1.0	1.0	1.0	3.0	5.5	1.0	1.5	7.0	5.0	1.0	1.0	1.0	1.0	7.0	7.0	4.0	5.5	1.0	1.5	7.0	7.0	1.0	1.0	7.0	7.0	4.5	5.5	distasteful
deep	1.0	2.0	1.0	2.0	1.0	1.5	4.5	4.0	2.0	2.5	1.0	1.0	2.5	1.5	1.0	1.0	7.0	1.0	2.5	3.5	1.0	2.0	4.0	1.0	1.0	2.0	1.0	1.0	1.5	3.5	shallow
active	1.0	2.5	1.0	1.0	1.0	1.0	4.0	5.0	2.5	4.0	1.0	1.0	1.0	1.0	7.0	3.5	4.0	2.0	4.5	4.5	1.0	2.0	1.0	2.0	2.5	1.0	1.0	1.5	4.0	5.0	passive

TABLE 3

RAW DATA FOR EVE BLACK

	LOVE		CHILD		MY DOCTOR		ME		MY JOB		MENTAL SICKNESS		MY MOTHER		PEACE OF MIND		FRAUD		MY SPOUSE		SELF-CONTROL		HATRED		MY FATHER		CON-FUSION		SEX		
	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II	
cold	4.0	1.0	1.0	1.0	7.0	7.0	7.0	7.0	1.0	1.0	1.0	1.0	1.0	1.5	7.0	5.5	6.5	6.0	2.5	1.0	3.5	2.0	6.0	6.0	7.0	6.5	4.0	1.0	1.0	1.0	hot
valuable	7.0	7.0	7.0	6.0	1.0	1.0	1.0	1.0	7.0	7.0	7.0	7.0	4.5	2.5	1.0	3.0	2.0	2.0	7.0	7.0	4.5	4.5	1.5	2.0	1.0	1.0	7.0	7.0	7.0	7.0	worthless
tense	1.0	1.0	1.0	1.5	7.0	7.0	7.0	7.0	4.0	1.0	2.5	1.0	4.0	2.0	7.0	6.5	6.0	6.0	4.0	2.0	2.0	2.0	6.5	5.0	1.0	2.0	1.5	1.0	1.0	1.0	relaxed
small	1.0	1.0	7.0	1.0	7.0	7.0	4.0	7.0	5.5	1.0	4.0	1.0	4.0	4.5	7.0	5.5	6.5	6.5	2.5	1.0	6.0	4.5	6.5	6.0	1.0	4.0	4.0	1.0	1.0	1.0	large
fast	7.0	1.0	4.0	3.5	1.0	1.0	1.0	2.0	7.0	7.0	1.0	1.0	1.0	1.0	1.0	1.0	1.5	1.5	7.0	6.5	2.0	1.5	1.0	1.5	4.0	1.5	4.5	1.5	4.0	5.5	slow
dirty	2.0	2.0	1.0	3.0	7.0	7.0	7.0	7.0	4.0	1.0	1.5	1.0	7.0	4.0	7.0	6.0	5.0	6.0	4.0	1.0	4.5	4.5	6.0	6.5	7.0	7.0	1.5	1.0	4.0	1.0	clean
weak	1.0	1.0	7.0	1.5	7.0	7.0	7.0	7.0	1.0	1.0	1.0	1.0	4.0	7.0	7.0	6.0	6.0	6.0	4.0	1.0	5.0	5.0	6.5	6.0	1.0	4.0	1.0	1.0	1.0	1.0	strong
tasty	7.0	7.0	7.0	6.0	1.5	1.0	1.0	1.0	7.0	7.0	7.0	7.0	7.0	5.5	1.0	1.5	2.0	2.0	5.5	7.0	3.5	4.5	2.0	2.0	1.0	1.0	4.0	7.0	7.0	7.0	distasteful
deep	7.0	7.0	7.0	2.5	1.0	1.0	1.0	1.0	7.0	7.0	4.0	7.0	7.0	1.5	1.0	2.0	1.5	2.0	4.0	7.0	2.0	1.5	1.5	2.0	7.0	2.0	4.0	4.0	7.0	7.0	shallow
active	7.0	7.0	7.0	7.0	1.0	1.0	1.0	1.0	7.0	7.0	7.0	7.0	1.0	1.0	1.0	1.0	1.5	1.5	7.0	7.0	4.5	4.0	1.5	1.5	1.0	1.0	4.0	7.0	7.0	7.0	passive

TABLE 4

RAW DATA FOR JANE

	LOVE		CHILD		MY DOCTOR		ME		MY JOB		MENTAL SICKNESS		MY MOTHER		PEACE OF MIND		FRAUD		MY SPOUSE		SELF-CONTROL		HATRED		MY FATHER		CON-FUSION		SEX		
	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II	
cold	4.0	6.0	4.0	6.5	4.0	6.5	4.0	5.5	4.0	6.0	4.0	4.0	4.5	4.5	4.0	6.5	1.0	2.0	4.0	6.0	4.0	6.0	1.0	1.5	6.0	6.5	4.0	4.5	4.0	6.5	hot
valuable	1.0	1.0	1.0	1.0	1.0	1.0	4.0	2.0	1.0	1.0	2.0	2.0	1.0	1.0	1.0	1.0	7.0	7.0	4.0	1.0	1.0	1.0	6.0	7.0	1.0	1.0	2.5	4.0	1.0	1.0	worthless
tense	7.0	7.0	5.0	5.0	7.0	7.0	2.0	1.0	1.5	2.0	1.0	1.0	1.5	2.0	7.0	7.0	1.5	2.0	7.0	7.0	7.0	6.5	1.0	1.5	1.5	2.0	1.0	1.5	7.0	7.0	relaxed
small	7.0	7.0	7.0	6.5	7.0	7.0	4.0	4.5	7.0	6.5	7.0	6.5	5.5	6.5	7.0	7.0	4.5	1.5	7.0	7.0	7.0	7.0	4.5	2.0	7.0	7.0	6.0	4.0	6.5	7.0	large
fast	7.0	6.5	3.0	1.5	1.0	2.0	4.5	2.0	7.0	7.0	7.0	6.0	3.5	2.0	7.0	6.0	1.5	2.0	3.0	2.5	7.0	6.0	7.0	4.0	2.0	2.0	1.5	2.0	4.0	4.0	slow
dirty	7.0	7.0	7.0	7.0	7.0	7.0	7.0	7.0	7.0	7.0	7.0	7.0	7.0	7.0	7.0	7.0	1.0	2.0	7.0	7.0	7.0	7.0	2.0	2.0	7.0	7.0	7.0	4.0	7.0	7.0	clean
weak	7.0	7.0	4.0	6.0	7.0	7.0	4.5	4.0	6.5	6.5	4.0	5.0	2.0	6.0	7.0	7.0	1.0	2.0	7.0	7.0	7.0	7.0	1.0	2.0	7.0	7.0	4.0	4.0	7.0	6.5	strong
tasty	1.0	1.0	1.0	1.0	1.0	1.0	3.5	2.0	4.5	2.0	5.0	4.0	2.0	1.0	1.0	1.0	7.0	7.0	1.0	1.0	1.0	1.0	7.0	7.0	1.0	1.0	7.0	5.5	1.0	1.0	distasteful
deep	1.0	1.0	1.0	2.0	1.0	1.5	2.0	3.0	1.0	1.5	1.0	1.5	1.0	2.0	1.0	1.5	1.0	2.0	2.0	2.0	1.0	1.5	1.0	2.0	1.0	1.5	1.0	1.5	4.0	1.0	shallow
active	2.5	1.5	1.0	1.0	1.0	1.0	2.0	1.5	4.5	2.0	1.0	1.0	1.0	1.0	5.5	2.5	6.0	4.5	4.0	1.0	2.5	2.0	2.5	2.0	1.0	1.0	1.5	2.5	4.0	1.5	passive

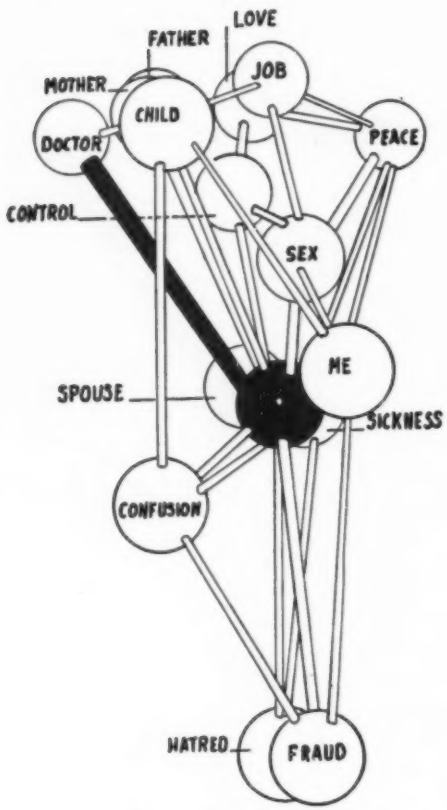


FIG. 1. EVE WHITE I

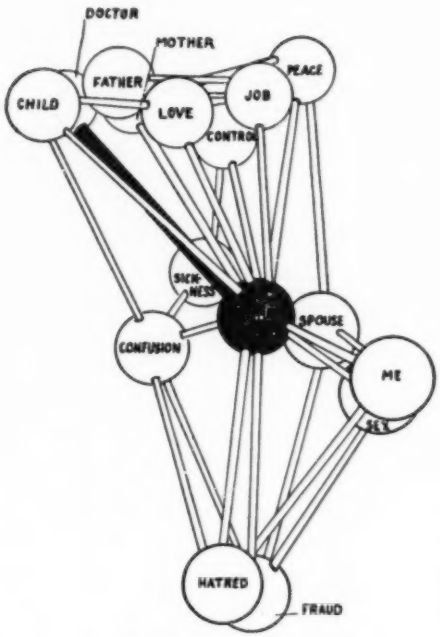


FIG. 2. EVE WHITE II

factory attitude toward herself. Here the usual societal "goods" are seen favorably—MY DOCTOR, MY FATHER, LOVE, SELF-CONTROL, PEACE OF MIND, and MY MOTHER are all *good* and *strong* whereas FRAUD, HATRED, and to some extent CONFUSION are *bad*. The chief evidence of disturbance in the personality is the fact that ME (the self concept) is considered a little *bad*, a little *passive*, and definitely *weak*. Substantiating evidence is the *weakness* of her CHILD, as she sees him (or her), and the essential meaninglessness to her of MY SPOUSE and SEX. Note also the wide evaluative separation between LOVE and SEX. In the interval between testings I and II ME and SEX become more *bad* and *passive* and simultaneously become almost identical in meaning to her—and note that her conceptions of LOVE (a good, strong thing) and SEX (a bad, weak thing like herself) have moved still further apart.

Eve Black

Semantic structures for Eve Black I and II are shown in Fig. 3 and 4. The most general characterization here would be that *Eve Black has achieved a violent kind of adjustment in which she perceives herself as literally perfect, but, to accomplish this break, her way of perceiving "the world" becomes completely disoriented from the norm*. The only exceptions to this dictum are MY DOCTOR and PEACE OF MIND, which maintain their *good* and *strong* characteristics, the latter, interestingly enough, also becoming *active* on II. But if Eve Black perceives herself as being *good*, then she also has to accept HATRED and FRAUD as positive values, since (we assume) she has strong hatreds and is socially fraudulent. So we find a tight, but very un-normal, favorable cluster of ME, MY DOCTOR, PEACE OF MIND, HATRED, and FRAUD. What are positive values for most people—CHILD, MY SPOUSE, MY JOB, LOVE, and SEX—are com-

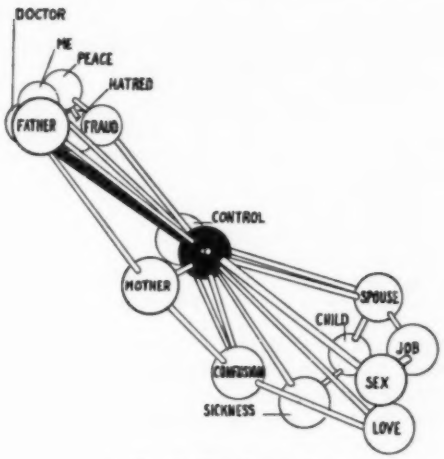


FIG. 3. EVE BLACK I

pletely rejected as *bad* and *passive*, and all of these except CHILD are also *weak* (this may be because CHILD was weak in Eve White and much of the change here is a simple "flip-flop" of meanings). Note that it is MOTHER in this personality that becomes relatively meaningless; FATHER, on the other hand, stays *good* but shifts completely from *strong* (in Eve White) to *weak*—possible implications of these familial identifications will be considered later. Note also that in this personality LOVE and SEX are closely identified, both as *bad*, *weak*, *passive* things.

Jane

Semantic structures for Jane I and II are shown in Fig. 5 and 6. The general characterization is that *Jane displays the most "healthy" meaning pattern, in which she accepts the usual evaluations of concepts by her*

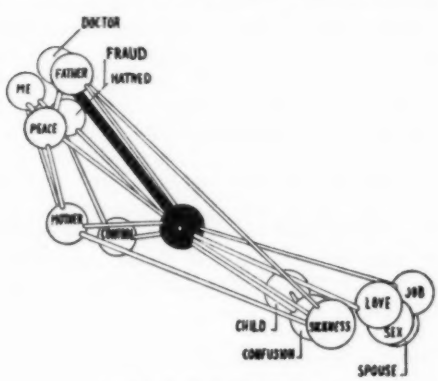


FIG. 4. EVE BLACK II

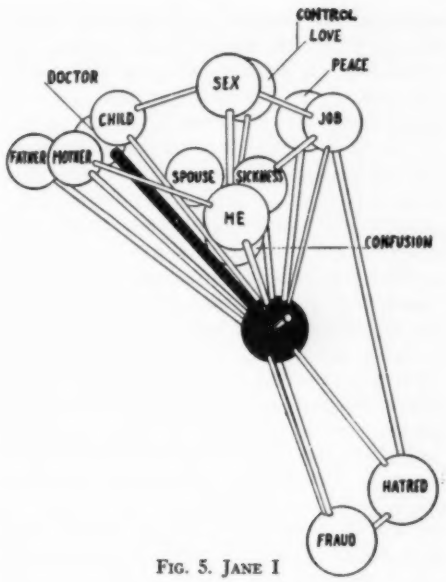


FIG. 5. JANE I

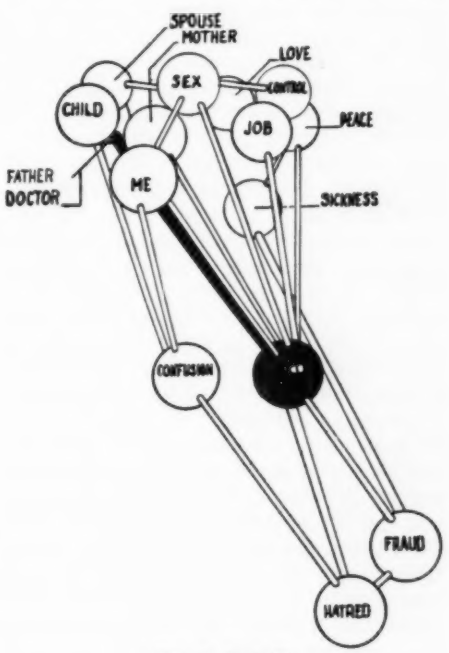


FIG. 6. JANE II

society yet still maintains a satisfactory evaluation of herself. MY FATHER, MY MOTHER, MY CHILD, and MY DOCTOR—most of the significant persons in her life—are seen as *good*, *strong*, and *active*. The major modes of behavior, PEACE OF MIND, LOVE, SELF-CONTROL, and MY JOB, are seen as equally *good* and *strong*, but *somewhat passive*—as if these ways of behaving and thinking were simply accepted without stress. The two socially agreed-upon evils, HATRED and FRAUD, are put in their proper places. The most significant characteristics of Jane's meaning system, however, are these: The self concept, ME, while still not *strong* (but not *weak*, either) is nearer the *good* and *active* directions of the semantic space; note also the close identification of ME and MENTAL SICKNESS, which here is *not* an unfavorable concept to her. Her attitude toward her husband, MY SPOUSE, is for the first time meaningful (unlike Eve White) and tending toward the *good*, *strong*, *active* directions, like the other significant persons (unlike Eve Black). And LOVE and SEX (quite unlike Eve White) are both favorable and quite closely identified. The changes from testings I to II are simply such as to strengthen the "healthy" pattern evident in the first view. ME becomes considerably more *good* and *active*; MY SPOUSE for the first time becomes completely identified connotatively with MY DOCTOR and MY FATHER (and loses its tie with CONFUSION); and LOVE and SEX become intimately identified with each other and close in meaning to SELF-CONTROL and PEACE OF MIND.

The thumbnail semantic sketches of each personality just given make it intuitively evident that the semantic differential does draw sharp distinctions between the three personalities inhabiting one nervous system. It is possible to demonstrate these distinctions quantitatively by intercorrelating *D* matrices between personalities and over time. If two of our models are generally similar in structure, such that large and small distances between concepts in one are reflected also in the other, then the *r* will be high. Table 5 gives these correlations. The first thing to note is that the correlation of each personality with itself (e.g., testings I and II) is regularly much higher than the correlation of that personality with any other personality (with the single exception of Eve White I and Jane I). This

TABLE 5

CORRELATIONS OF *D*-MATRICES BETWEEN PERSONALITIES AND OVER TIME

	White I	White II	Black I	Black II	Jane I	Jane II
White I	—					
White II	.73	—				
Black I	−.06	—	—			
Black II	—	−.02	.86	—		
Jane I	.73	—	−.26	—	—	
Jane II	—	.53	—	−.08	.92	—

is quantitative justification for the statement that the semantic differential does differentiate between the several personalities of this woman. Whether it differentiates in a valid way is a matter that can be judged only by relating our analysis to the detailed case history material available elsewhere (3).

Another important thing to note about these correlations is that Eve White and Jane (the two "socialized" personalities) are fairly highly correlated whereas the correlations of Eve Black with the other two are definitely low, even negative. In other words, Eve Black is clearly the most deviant and disordered personality. Finally, it should be noted that these three personalities differ somewhat in their stability, as indexed by the I/II correlations, Eve White being the least stable and Jane the most.

CHANGES IN MEANING OF SPECIFIC CONCEPTS

As noted earlier, the meaning of a specific concept is operationally defined as its profile against the differential (e.g., its position in semantic space). Change in meaning between testings I and II can be measured directly by the *D* between I and II profiles for the same concept (e.g., between the paired columns in Tables 2–4). These *D* values are given under "Within Personalities, Between Testings" in Table 6. Changes in meaning between personalities for the same concepts can be measured directly by the *D* between profiles for the same concept but as judged in two different personalities; these *D* values are given under "Between Personalities, Within Testings" in Table 6.

Semantic Stability Through Time

In general, although the differences are not great, Eve Black is the least stable personality

TABLE 6

CHANGES IN MEANING OF SPECIFIC CONCEPTS

Concept	Within Personalities, Between Testings			Between Personalities, Within Testings					
	$D_{I\text{-}II}$ Eve White	$D_{I\text{-}II}$ Eve Black	$D_{I\text{-}II}$ Jane	$D_{W\text{-}B}$		$D_{W\text{-}J}$		$D_{B\text{-}J}$	
				I	II	I	II	I	II
LOVE	.42	.67	.23	1.58	1.44	.35	.57	1.62	1.81
CHILD	.71	.96	.37	1.65	1.40	.68	.54	1.47	1.41
MY DOCTOR	.15	.05	.27	.23	.23	.28	.25	.30	.12
ME	.36	.32	.42	1.21	1.40	*.60*	*.88*	.83	.77
MY JOB	.27	.62	.42	1.30	1.43	.49	.43	1.19	1.54
MENTAL SICKNESS	.45	.45	.19	1.24	1.38	.40	.32	1.30	1.47
MY MOTHER	.54	.78	.46	.71	.78	.66	.23	1.02	.68
PEACE OF MIND	.40	.35	.41	.86	.66	.21	.28	.81	.61
FRAUD	.64	.12	.40	1.46	1.35	.73	.34	1.29	1.22
MY SPOUSE	.30	.62	.47	.67	.96	*.61*	*.89*	1.04	1.75
SELF-CONTROL	.32	.25	.24	.78	.80	.34	.39	.92	1.01
HATRED	.51	.19	.44	1.54	1.31	.51	.23	1.37	1.19
MY FATHER	.53	.71	.09	1.06	.60	.25	.43	1.06	.43
CONFUSION	.64	.67	.44	.86	.96	.71	.42	.98	.88
SEX	.62	.34	.47	1.10	.62	*.63*	*1.20*	1.45	1.76

Note.—Numbers in italics indicate concepts that serve best to characterize differences between Eve White and Jane.

through time and Jane the most stable (cf., columns 2–4 in Table 6). The concept-by-concept data thus confirm the stability of Jane as shown in the structural data given above. For Eve White the most unstable or labile concepts are CHILD, MY MOTHER, FRAUD, HATRED, MY FATHER, CONFUSION, and SEX. For EVE Black the most labile concepts are LOVE, CHILD, MY JOB, MY MOTHER, MY SPOUSE, MY FATHER, and CONFUSION. For Jane the most unstable notions are ME, MY JOB, MY MOTHER, MY SPOUSE, HATRED, CONFUSION, and SEX. We note that the family constellation—mother, father, spouse, child—tend to be more susceptible to change through time within these personalities, but that the self concept stays relatively constant within personalities (e.g., the location of the self concept, in a sense, defines these three personalities). HATRED, SEX, and CONFUSION also seem to be points of stress.

Semantic Stability Between Personalities

The data given in columns 5–10 in Table 6 make it clear that concept meanings in general shift more between Eve Black and the other two than between Jane and Eve White, this again substantiating the over-all correlations between total structures. The only concept that remains strictly constant in meaning through the personality changes in this woman is MY DOCTOR, although PEACE OF MIND and CONFUSION show greater stability than most others. It is interesting to note which concepts serve best to characterize the differences between Eve White and Jane (Eve Black shows gross differences on almost all concepts). These two "socialized" personalities differ from one another chiefly on ME, MY SPOUSE, and SEX, and these differences are increasing in magnitude through time. This clearly suggests this woman's sexual life as a core problem, Eve White being highly critical of all three concepts and Jane accepting them as positive values. It is also interesting to note in this connection that semantic differences between Eve White and Jane on CHILD, MY MOTHER, FRAUD, HATRED, and CONFUSION are decreasing through time.

INTERPRETATIONS OF THESE SEMANTIC DATA

The analyses of these personalities and their changes given so far have been descriptive rather than interpretive for the most part. In a sense, we have merely put into words what this woman herself, in her several personalities, has indicated by her check marks. The treatment of semantic differential data, from the patterns of check marks to construction of the models shown in Fig. 1–6, is completely objective, and any investigator starting from the same data and following the rules would have to end up with the same pictures we have.

Making interpretations and predictions about this case on a "blind" basis is another matter entirely. In this section we go far beyond the objective data, and we are consequently much less confident about our statements. For one thing, neither of the writers is an experienced clinician—certainly not experienced with respect to the dynamics and characteristics of multiple personality. For another thing, we do not know at what stage in therapy our two testings were made, and interpretation would certainly vary greatly in terms of such information. It should also be pointed out that in the ordinary use of the semantic differential as a clinical tool (as compared with a blind analysis) many other sources of information would be available to support certain alternative interpretations and render others farcical. Let it be understood, then, that what follows is a flight into

conjecture, in contrast with the preceding, factual reporting of semantic data.

Interpretive Descriptions of the Three Personalities

Eve White is the woman who is simultaneously most in contact with social reality and under the greatest emotional stress. She is aware of both the demands of society and her own inadequacies in meeting them. She sees herself as a passive weakling and is also consciously aware of the discord in her sexual life, drawing increasingly sharp distinctions between LOVE as an idealized notion and SEX as a crude reality. She maintains the greatest diversity among the meanings of various concepts. She is concerned and ambivalent about her CHILD, but apparently is *not* aware of her own ambivalent attitudes toward her MOTHER—and seems to become more resistant to this by testing II. Those psychoanalytically inclined may wish to identify Eve White with dominance of the *superego*: certainly, the superego seems to view the world through the eyes of Eve White, accepting the mores or values of others (particularly her mother) but continuously criticizing and punishing herself. If this case came to the psychotherapists with a voluntary, self-initiated plea for help, then it seems likely that Eve White was dominant at the time.

Eve Black is clearly the most out of contact with social reality and simultaneously the most self-assured. To rhapsodize, Eve Black finds PEACE OF MIND through close identification with a God-like therapist (MY DOCTOR, probably a father symbol for her), accepting her HATRED and FRAUD as perfectly legitimate aspects of the God-like role. Naturally, she sees herself as a dominant, active wonder-woman and is in no way self-critical. She is probably unaware of her family situation. Those psychoanalytically inclined could say that the *id* looks out at the world through the eyes of Eve Black. Like a completely selfish infant, this personality is entirely oriented around the assumption of its own perfection. Actually, Eve Black seems to be more harmonious with the Adlerian than with the Freudian model, since personal perfection is apparently the demand acceded to rather than sexuality. If the case was committed to an institution, it seems likely that this personality was the reason for commitment.

Jane is the most puzzling of the three personalities, and our interpretation will have to depend upon assumptions about the stage of treatment (see below). Superficially, Jane is a very healthy personality: "all's well with the world, and day by day I'm getting better and better." Thus we find all the people in her life perceived as *good* and *strong* and *active* and all the socially approved modes of action perceived as *good* and *strong* and *passive*; SEX is LOVE-ly, her SPOUSE is becoming more like the noble DOCTOR all the time, and she is coming to perceive herself even, as a pleasant and reasonably active (if somewhat weak and submissive) person. But all this is a little too rosy, a little too pat. We note that Jane is becoming more and more "simple-minded"—all of her judgments tending to fall along a single factor of *good-strong* vs. *bad-weak*—which makes the Jane II model the most restricted and undiversified of all. Those psychoanalytically inclined may wish to view this personality as representing dominance of a self-deceptive *ego* which has woven a web of repression as to the state of reality; or, they may wish to view Jane as an essentially strong, healthy, and improving ego-dominated personality. In any case, we doubt if Jane would have either come for therapy or have been institutionalized—as such.

Possible Dynamisms Operating in the Case

Identification mechanisms. We say the patient "identifies" with some other person when her meaning of herself, ME, is semantically close to her meaning of the other person; e.g., if she sees her father as a kind, active, relaxed, etc. person and describes herself in the same terms, we infer identification. However, the pattern of identifications displayed by this patient seems unusual. Only in Eve Black, the obviously disoriented personality in terms of her values, is there clear differential identification—with her FATHER (and this may reflect the semantic tie-up between FATHER and MY DOCTOR). Jane shows some slight tendency toward closer identification with MOTHER, but it is not close. Eve White shows none with either parent. The fact that identification with FATHER (and MY DOCTOR) in Eve Black is accompanied by rejection of MOTHER to meaninglessness is suggestive of an underlying conflict in identifications. Note also, in this connection, that in Eve Black I the

ascendancy of ME to the *good, strong, active* position is accompanied by making FATHER *weak*—as if she were taking over her father's role and putting her mother in her own previous place. And, interestingly enough, the concept SELF-CONTROL suffers the same fate as MOTHER. This picture of Eve Black is certainly suggestive of an *Electra complex* as the underlying dynamism. In "real" life, her MOTHER is or was the dominant, threatening figure—moralizing, demanding standards and SELF-CONTROL—and in Eve Black this woman escapes the pressure by rendering both MOTHER and SELF-CONTROL meaningless and simultaneously identifying with and taking her FATHER'S place (via the therapist). Suggestive evidence may be found in Tables 2–4: MOTHER is consistently *colder* than FATHER and usually more *tense* and *fast* (e.g., Factor 3). Identification of the self with the therapist in Eve Black is perfect, of course. The concept MY DOCTOR is the only personal concept to show perfect stability both between personalities and through time. The patient thus displays what might be called maximal *positive transference* in all three personalities; there is no sign of any negative transference at either testing, which may be indicative of the stage of therapy (e.g., early).

Significance of the patient's sexual life. Although Jane shows a rosy acceptance of normal sexual patterning—with LOVE and SEX linked and passively favorable, Eve White clearly displays awareness of a basic conflict in this area—SEX is early somewhat more distasteful than LOVE and becomes distinctly distasteful and dirty by testing II. In Eve White also we find ME and MY SPOUSE becoming linked with SEX in this unpleasant location. Eve Black, on the other hand, rejects both SEX and LOVE—but closely links them in her thinking. If we were to relate these facts with the Electra situation described above, the interpretation would be that her persisting conflict with her mother and attempts to identify with her father make it impossible for her to experience normal sexual satisfactions with her husband and to carry out the normal mother-wife-home role. Eve White is aware of this, in a sense, but Jane clearly is not. The concept MY JOB is interesting in this connection: its persistent linkages with LOVE, PEACE OF MIND, and SELF-CONTROL in the two "socialized" personalities, coupled with its linkage with SEX and MY SPOUSE in Eve Black, clearly suggests to us that this woman is interpreting MY JOB in the sense of "my job as a mother, wife, and homebuilder" rather than in terms of her outside work (which we understand she has). In any case, there is clear evidence of involvement with her sex life as a major problem, and this may have been the presenting problem when she began therapy.

Repression and amnesia. Knowing that we are dealing with a case of multiple personality—usually characterized by complete dissociation between states—it is interesting to speculate on what meanings are repressed in the several personalities. It will be recalled that, operationally, meaninglessness of a concept is defined by its closeness to the origin (the solid balls in the models). This is probably to be interpreted as "connotative deadness" or "damping of affect" with respect to the concept involved. Within the matrix of our instrument, however, there is another way in which repression or amnesia may show up, and that is via a complete shift in meaning of the concept being judged (e.g., CHILD may shift from the personal reference of "my own youngster" to "children-in-general"). Looking back at the semantic data with these points in mind, we hazard the following guesses: Eve White probably has the best contact with reality and may not be amnesic at all (except for the other personalities); Eve Black may be amnesic for her mother and her own role as a mother and wife; Jane in Pollyanna fashion may be amnesic for her own problems, e.g., MENTAL SICKNESS and CONFUSION, and the indiscriminate way in which she lumps all socially favorable concepts at least suggests that she is judging CHILD, MOTHER, etc. in the abstract rather than as MY CHILD, MY MOTHER, and so forth.

Interpretation I—Assumed Early Stage of Therapy

The "original" personality, in the sense of being most characteristic of the woman her friends and relations knew, was Jane. The first testing of this personality shows a relatively weak ME that is associated with MENTAL SICKNESS; i.e., she was dimly aware of her own inadequacies but was striving to maintain a rigid acceptance of the real world and maintain an adequate home life. The people about her, with the exception of her husband,

were seen as strong and active (perhaps threatening) in relation to herself, and her love life was regarded as a sort of deliberate, controlled duty. She was completely unaware of her (repressed) emotional ambivalence toward her mother, husband, and child. The things being hidden in this personality, and providing the force behind the eventual split, were (*a*) her Electra complex, (*b*) her repugnance for sexual relations with her husband, and (*c*) her ambivalent attitude toward herself. We suspect that she had a position in society that demanded "good front."

We must assume strong and about equal pressures toward solving the Electra complex, (*a*) by identifying with FATHER and asserting the self (id?), and (*b*) toward solving it by identifying with MOTHER and devaluing herself (superego?). This produces a two-way split away from the Jane pattern, one into Eve Black where selfish needs for superiority and playing the father role are achieved and another into Eve White where societal needs for submission and playing the mother role are achieved. This split, and the subsequent availability of the other roles, allows Jane to shift toward the "sweetness and light" view of the world, and this is clearly demonstrated by the changes between Jane I and Jane II. Eve White continues to become more simply and rigidly self-critical and Eve Black continues to become more simply and rigidly self-satisfied.

Assuming successful therapy is possible—which seems questionable—it will involve less and less time being spent in being Jane and Eve Black and a consequent shift into Eve White, where better contact and differentiation seems to be maintained. But here it will be necessary to bring Eve White to understand the reason for her depression, the role of her ambivalence toward her mother in her problem—which shows no signs of happening yet—and thence a gradual restructuring in which ME becomes more favorable, along with SPOUSE and SEX, and identifications with FATHER and MOTHER are reassigned. This will probably involve a period of negative transference, with MY DOCTOR and MOTHER becoming closely identified and being temporarily shifted to *bad*, *strong*, and *active* directions of the semantic space. In other words, successful treatment will mean increasing time spent in, and a gradual restructuring of, Eve White to the point where it incorporates what is now Jane, but with a realignment of significant persons. On the other hand, if this woman is in a mental institution and remains there, it seems likely that Eve Black will become the dominant house she lives in. In either case, it is probable that Jane will appear less and less.

Interpretation II—Assumed Late Stage of Therapy

If we assume that we are seeing the terminal stages of therapy with a case of this sort, then a quite different interpretation is necessary. The difference in interpretations hinges upon Jane, either as a deceptive and vanishing original personality (interpretation I) or as an increasingly healthy and augmenting personality (interpretation II). In the latter case we would assume that Eve White had been the "original" personality as people knew her—a socially acceptable wife and mother, but one laden with conflicts, anxieties, and self-criticism. The split in this case—into one personality in which the self-criticism completely disappears via irrationality (Eve Black) and another in which self-criticism vanishes via rationality (Jane)—seems less sensible, however. Jane seems unnecessary at this stage and really should have developed out of Eve White rather than being contemporaneous. If we assume there was a split in any case, Jane is clearly the most healthy personality, since LOVE and SEX are identified, the world is viewed in acceptable fashion, and the self concept is becoming more favorable all the time. The prediction here would be increasing time spent in Jane and less in the others. The stumbling blocks in the way of this interpretation are (*a*) the lack of any realignment of the system of parental identifications, and (*b*) the fact that Jane is becoming *less* diversified semantically (more "simple-minded") rather than the reverse. This second interpretation was actually the one we first adopted—because of the superficial "healthiness" of Jane—but consideration of all the evidence seems to favor the first interpretation. However, it should be noted that *if* this case is near the end of successful therapy, Jane is the only personality that combines both a normal view of the world and reasonable (increasing) acceptance of the self.

Interpretation III—Combination of I and II

It is possible to combine interpretations I and II by assuming that Jane is both the original personality which broke apart and the terminal personality which is being developed out of therapy. In this case, the early development of the case, probably in childhood, would be the same as that given under interpretation I—the conflicting parental identifications (id and superego determined respectively) were of about equal strength and finally became too intense to be contained within the self-deceptive personality organization of Jane. During the middle course of the case, when therapy was undertaken, we thus find all three personalities oscillating, temporary dominance of the mother and wife role being represented by Eve White, temporary dominance of the self-gratifying father role being represented by Eve Black, and temporary dominance of the face-saving, problem-solving ego being represented by Jane. Intentionally or unintentionally, the effect of therapy may be to strengthen the self-deceptive organization of Jane without resolving the underlying conflicts dramatized by Eve White and Eve Black. The oversimplified, Pollyanna-like ways of perceiving herself as *good* along with all the other significant persons in her life yields a superficially happy person who views the world in an acceptable, if rigidly stereotyped, fashion. If the present combined interpretation approximates the actual situation, then we feel compelled to predict another breakdown at some later period in this person's life. In other words, the effect of therapy (whatever type it may have been) seems to have been further to strengthen the self-deceptive original organization of Jane, while making this personality even more rigid and insensitive to subtle differences in meaning and without resolving the underlying conflicts which created the original disturbance.

What Price Therapy?

It is impossible to tell from our semantic data whether the increasing simplification in structure characteristic of all three personalities is due to therapy itself or is happening despite therapy. However, a number of specialists in psychotherapy have from time to time expressed concern over the "hidden" effects of therapy even in so-called successful cases—particularly reduction in initiative, creativeness, and flexibility of the patient. Certainly in the present case we are witnessing an over-all reduction in differentiation of meanings. If overt behavior is in considerable part determined by meanings, as we believe it is, then we must expect Jane (if she is the terminal personality) to be now even less capable of behaving differentially to her mother, father, spouse, and child—they are all essentially undifferentiated "strong-active-goodness" to her. This would also be true, but to a lesser extent, of Eve White, although here we would assume an earlier stage of therapy and hence a possibility of secondary elaboration of semantic diversity under sensitive therapy.

Is rigidity of this sort a necessary price of therapy? In striving to achieve the goals of societal acceptability and invididual happiness, does the therapist have to sacrifice the richness, individuality, and subtler adjustiveness of the patient? These are serious questions raised by the data of this single case—but not answered by them, of course. From the larger sample of cases we are presently working on, better answers may be forthcoming, but the cases are generally less severe. One other interesting phenomenon in the present cases should be mentioned: Despite the gross changes in meaning of concepts in the several personalities, and the over-all reduction in diversity, the semantic judgmental frame of reference remains constant. In other words, all three of the personalities in this woman utilize semantic scales in the same ways—the correlations between scales are the same for all three personalities and reduction in diversity in all of them is accomplished by a coalescence of *good*, *strong*, and *active* into a single evaluative dimension. Thus it would appear that the level of scale meanings is below that at which concepts vary, and common to all three personalities.

REFERENCES

1. OSGOOD, C. E. The nature and measurement of meaning. *Psychol. Bull.*, 1952, **49**, 192–237.
2. OSGOOD, C. E., & SUCI, G. J. A measure of relation determined by both mean difference and profile information. *Psychol. Bull.*, 1952, **49**, 251–262.
3. THIGPEN, C. H., & CLECKLEY, H. A case of multiple personality. *J. abnorm. soc. Psychol.*, 1954, **49**, 135–151.

Received December 1, 1953.

CRITIQUE AND NOTES

EFFECT OF INCREASING THE NUMBER OF RORSCHACH RESPONSES ON *SUM C* AND *M*: A NOTE ON FISKE AND BAUGHMAN'S STUDY

CONAN KORNETSKY AND DONALD L. GERARD

Laboratory for Socio Environmental Studies, National Institute of Mental Health, Bethesda, Maryland

FISKE AND BAUGHMAN in their recently published study inquired "... to what extent are the frequencies of responses in the several scoring categories related to the total number of responses?" (2, p. 25). Their data indicated that there was a positive relationship between the number of responses (*R*) and a number of clinically important Rorschach determinants. They state, "While it is clear that frequencies in most categories are related to the total number of responses, we cannot state, on the basis of the present data, which influences the other" (2, p. 30). This question which Fiske and Baughman ask may be partially answered as follows. If additional responses are elicited from a group of subjects (*S*s) who spontaneously produced comparatively few responses on the Rorschach so that their total number of responses will equal those of a group who initially produced a larger number of responses, the influence of *R* on various scoring categories may be determined. Thus, it would be reasonable to expect that if the sum of the color responses (*Sum C*) and the human fantasy responses (*M*)[1] are a function of *R*, there would be no statistically significant difference between the amount of *Sum C* and *M* in these two groups when the number of responses are equated.

METHOD

Data which could be utilized for this purpose were obtained in a current study of teen-age opiate addicts. A procedure similar to that used by Jernigan (3) was employed. The Rorschach was first given in the usual manner. After this, but before the inquiry, the cards were again presented to the *S* with the instruction to tell again everything that he sees, but not to give any responses that he had given in the previous administration.

Twenty-two *S*s were studied in this manner. They were divided into two groups: Group A consisted of *S*s who had 18 or more responses on their first Rorschach administration; group B consisted of *S*s who had less than 18 responses on their first Rorschach (mean: 15.3, range: 13–16) but who were able to add sufficient new responses on their second Rorschach so that the total number of responses for Rorschach 1 and 2 was over 18.[2] For group B the determinants for Rorschach 1 and 2 were pooled, i.e., *Sum C* consisted of *Sum C* on Rorschach 1 plus *Sum C* on Rorschach 2. The distributions of the total number of responses for group A and group B were not significantly different (group A, mean: 25.0, range: 20–31; group B, mean 23.4, range: 19–30).

The *S*s were dichotomized at the medians of the total distribution of *Sum C* (median: 1.0) and *M* (median: 1.3). Fisher's "exact" test of the probability of obtaining a set of frequencies in a 2×2 table was used (1, p. 84) to evaluate the significance of the difference between *S*s above and below the dichotomy for the two groups.

RESULTS

The data for *Sum C* and *M* are summarized in Table 1 and 2 respectively. The probability that these obtained sets of frequencies would occur by chance are $< .03$ for *Sum C* and $< .06$ for *M*.

TABLE 1

AMOUNT OF *Sum C* DICHOTOMIZED AT 1.0 FOR GROUPS A AND B

Group	*Sum C* > 1.0	*Sum C* < 1.0
A	8	3
B	1	6

TABLE 2

AMOUNT OF *M* DICHOTOMIZED AT 1.3 FOR GROUPS A AND B

Group	*M* > 1.3	*M* < 1.3
A	7	4
B	1	6

DISCUSSION

These results indicate that equating the number of responses will not equate the frequency of color or human movement-determined responses. In terms of the problem raised by Fiske and Baughman, the results suggest that the frequency of color and human movement-determined responses is

[1] In the present study *Sum C* and *M* were selected because of: (*a*) their clinical importance, (*b*) their importance in a study now in preparation by the senior author, and (*c*) they appeared frequently enough in the small sample used so that statistical comparisons could be made.

[2] Four *S*s were discarded because their total number of responses for Rorschach 1 and 2 was not greater than 18.

probably not due to the influence of the number of responses as an independent variable.

The *Ss* who initially proffered the smaller number of responses were a constricted nonproductive group. The failure of this group to increase the number of affective responses so that it would equal a more productive group when additional percepts were forced may be the consequence of their personality structure. Thus a study that used proportions in comparing the amount of affective responses between a constricted nonproductive group and a nonconstricted productive group may negate clinically meaningful differences. On the other hand, when nonconstricted productive individuals add additional percepts, they may continue to use affective responses in the same proportion as they previously did, so that in a comparison between two productive groups the use of proportions may be justified.

REFERENCES

1. EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
2. FISKE, D. W., & BAUGHMAN, E. E. Relationship between Rorschach scoring categories and the total number of responses. *J. abnorm. soc. Psychol.*, 1953, **48**, 25–32.
3. JERNIGAN, J. A. A Rorschach study of normal and psychotic subjects in a situation of stress. Unpublished doctor's dissertation, Univer. of Kentucky, 1951.

Received July 17, 1953.

ADEQUACY OF THE REPORTING-TEST MEASURES FOR DETECTING REACTIONS TO FRUSTRATION[1]

W. BRUCE DOUGLASS
University of California, Los Angeles
AND KENNETH B. BROWN
University of Missouri

RECENTLY McKinney *et al.* (1) reported on a paper and pencil test designed to measure changes in performance due to frustration in a group situation. This device, called the Reporting Test, consists of three similar series of diagrams having varying numbers of three types of symbols which the subject (*S*) is required to count. In the first series, Condition I (C I), *S* is allowed to work at his own pace for eight minutes. In Condition II (C II), the frustrating situation, he is given an impossible task of completing a designated number of diagrams every 30 seconds, for two eight-minute periods. To emphasize his failure, *S* is required to mark an "F" for every problem not completed in each 30-second period. The final condition (C III) is an eight-minute postfrustration period in which *S* again works at his own speed.

Scores for each of the conditions on the Reporting Test can be in terms of problems attempted, net correct, errors, and percentage correct. The differences in scores between the nonfrustrating condition, C I, and the frustrating condition, C II, presumably are indicative of proneness to frustration. McKinney *et al.* (1) present evidence for a group of 196 Navy radar trainees that shows that under C II as compared with C I *Ss* made significantly more errors and attempted more problems, but had fewer net correct and a smaller percentage correct on the average. "However, not all individuals reacted in the same way to experimental frustration: the majority speeded up performance at the expense of accuracy, but a smaller group showed stable performance without increase of speed or errors" (1, p. 322). Similar results were found when the test was repeated on 98 college students.

While the investigators do not indicate what they consider the most appropriate method of scoring, they do state that the percentage correct score ". . . appears to be the most promising measure of efficiency . . ." (1, p. 320).

The present study was designed to investigate the reliability and, indirectly, the validity of the four scoring methods for relatively unselected college students. It is based on the assumption, supported by the results cited, that the frustrating situation of C II should produce different changes in performance scores for different individuals from the scores earned under C I. That is, a frustrating situation would be expected to have more effect on some *Ss* than on others. For an adequate measure, one would then expect (*a*) high correlations among the various parts of the test when the conditions are fairly constant and (*b*) relatively low correlations when the frustrating condition is introduced.

[1] This article is based upon the senior author's master's thesis (2) done under the junior author's direction, and submitted to the department of psychology, University of Missouri.

The above reasoning provided a basis for a fairly simple method of investigating the four scoring techniques used on the Reporting Test. This was done by setting up a nonfrustrated (NF) group, which was administered the three parts of the test without the frustrating condition. Another group of *Ss*, the frustrated (F) group, was given the test according to the standard procedure. Following this, intercorrelations were obtained among the three parts of the test for each group on each measure (problems attempted, net correct, errors, and percentage correct). The NF-group correlations would provide measures of reliability. Then a comparison of the F-group correlations with the paired NF-group correlations would give an indication of which types of scores significantly reflected the differential changes that the frustrating condition was expected to produce.

The NF group consisted of 101 university students (74 males and 27 females) from two general psychology classes. The F group also had 101 *Ss* (65 males and 36 females) from two other general psychology classes.

The NF group was given the same instructions at the beginning as were the *Ss* of the F group. However, after the first eight-minute period (C I), NF-group *Ss* were stopped and told that *E* was interested in determining the relative difficulty of the different parts of the test. They were then instructed to begin working on the C II portion of the test. This was done also for the second eight-minute section of C II and for C III.

A total of 28 product-moment correlations was calculated for each group, and the significance of differences between paired F- and NF-group r's was tested by the z transformation technique. However, only a summary of the results will be given here.

All the obtained r's for the two groups were significantly positive at beyond the .01 level of significance (r of .25 required for 99 *df*). The correlations for the NF group may be considered measures of reliability of the four types of scores used. However, few of the r's obtained for the NF group were high. For both error and percentage correct scores, none of the intercorrelations among various parts of the test was above .68. For problems attempted and net correct the r's for C I and C II or C III faired no better (highest r was .61). However, when the two sections of C II (independently and combined) were correlated with C III, the r's for both measures were appreciably higher, ranging between .78 and .87.

These results for the NF group suggest that for all four measures scores obtained on the prefrustration period may be too unreliable to be used in individual analysis. That is, *Ss* of the NF group did not reach fairly stable relative performance levels, as shown by problems attempted and net correct, until after the first eight-minute period. This further suggests that an unscored "warm-up" period would be beneficial, so that an individual's performance under frustration could be compared with his stabilized performance without frustration.

When the r's for the F group were compared with the corresponding r's for the NF group, few of the F-group r's were found to be significantly lower. None would be expected to be significantly higher, and this was found to be the case. When the group differences in correlations between C I and C II were investigated, only those for the percentage correct scores were found to be significantly different in the expected direction (NF r = .66, F r = .46; p = .05). However, this measure faired poorly in other comparisons (C II vs. C III and C I vs. C III) as indicated below. Since the results given previously suggest that it is relatively unreliable, the over-all results do not appear to indicate that the percentage correct score is a dependable measure of individual performance changes under frustration.

When the differences between the paired F-group and NF-group correlations for the C II and C III periods were treated statistically, those for the F group were found to be significantly lower (p < .01) for both problems attempted and net correct scores. The respective r's obtained are as follows: For problems attempted, NF r = .87 and F r = .60; for net correct, NF r = .86 and F r = .70. Further study is needed to clarify the significance of changes in performance from frustrating to postfrustration conditions. Casual inspection of the test data suggests that some *Ss* showed a carryover of frustration reactions, others recovered quickly, others had little change from C I under either condition, while still others evidenced little change under C II but manifested it during the following period. While our results seem to indicate that dependable differential changes occur, their nature and meaning in terms of personality variables need further investigation.

Perhaps in part because of the low reliability of the error scores, no significant differences in the correlations of the F and NF groups were found on this measure. There were as many F-group r's that were higher than were lower than the comparable NF-group values. These results, along with the previously indicated low reliability of this measure, suggest that errors as well as percentage correct are not an adequate indicator of reactions to frustration, at least as the test is presently designed.

While there is considerable indication of face validity of the Reporting Test in observations of *Ss* taking it, in their introspective comments on it, as well as in the design of the frustrating situation itself, the present study suggests that refinement in scoring and construction would improve its usefulness for determining frustration tolerance.

REFERENCES

1. MCKINNEY, F., STROTHER, G. B., HINES, RUTH R., & ALLEE, RUTH A. Experimental frustration in a

group test situation. *J. abnorm. soc. Psychol.*, 1951, **46**, 316–323.
2. DOUGLASS, W. B. An investigation of various scoring methods for the McKinney Reporting Test. Unpublished master's thesis, Univer. of Missouri, 1952.

Received June 22, 1953.

THE STABILITY OF AUTOKINETIC JUDGMENTS[1]

J. H. ROHRER, S. H. BARON, E. L. HOFFMAN, AND D. V. SWANDER
Urban Life Research Institute, Tulane University

A MAJOR criterion used in a series of studies concerned with problems of personality structure and interpersonal interactions, carried out by the writers, has been rate of convergence to a common social norm established to perceived autokinetic light movement. Generalizations developed from this experimental model of a social situation are limited by the stability, in time, of the norms so established. Sherif (8) in his classic experiment demonstrated that such norms were stable when tested after a period of one week. Bovard (2) reported stability of norms after 28 days. This report deals with the stability of socially affected judgments, established to an autokinetic producing light *S*, after an elapsed time interval of one year.

METHOD

Subjects. The stability-of-norms test was carried out with 46 observers (*O*s) who had established social norms one year prior. All *O*s were undergraduate male students at Tulane and all were enrolled in either Army, Navy, or Air Force ROTC training units. None of the *O*s was aware of the purpose of the study and none had experienced the autokinetic-producing situation in the interval between test-retest.

The experimental situation and room were identical for the norm formation and retesting sessions. The experimental situation has been described in detail elsewhere (4). The apparatus also has been described elsewhere (7).

Experimental sessions. Three sessions were utilized: "T" an individual training session; "X" a social norm formation session; and, "Y" an individual test of norm stability. The method of training utilized in session T has been described elsewhere (4). Briefly, *O*s were trained in individual sessions using actual movement of a pinpoint of light within a circular luminous framework. The *O*s reported their judgments of the extent of movement after each exposure. Fifty exposures were made in this session. The actual distance that the light moved was determined by the experimenter (*E*) at the outset of the session and remained constant throughout the session, although the direction of the movement was systematically varied on a plane perpendicular to the *O* as permitted by the apparatus (7).

About one hour following the training session, the norm formation session occurred. In this session the luminous framework was absent, the pinpoint of light remained stationary, and the auditory cues from the apparatus were present. In session X the *O*s were tested in pairs, one member had been trained in session T at a distance of 2 in., the other member had been tested at a distance of 8 in. During session X a verbal report of the perceived extent of light movement was obtained from the *O*s after each exposure. Fifty presentations were also made in this session.

In session Y, which took place approximately 12 months following session X, each *O* was tested individually and judgments were given following each exposure. Fifty exposures were given in this session also. Thus, in session Y the conditions were identical with those used in session X, with the exception that each *O* was tested individually.

Scoring. The scores were in terms of inches of movement reported by the *O*. The median of each *O*'s 50 judgments, for each session, was computed. It is this median that constitutes the *O*'s norm or score for each session. References to "mean score" indicate the mean of medians.

RESULTS

The data for the *O*s trained with a 2-in. light movement were analyzed separately from the data obtained from the 8-in. training group. This was done because the means of the T scores for the groups differ significantly—which was to be expected—and the groups were heterogeneous with respect to variances. The results of an *F* test provide evidence for rejecting the hypothesis of homogeneous variance at the .01 level of confidence. Furthermore, it was not feasible to adjust the autokinetic X scores to account for the differences in T scores because an analysis of the regression data revealed that one regression line

[1] This study is one of a series carried out under ONR contract Nonr-475(01) with Tulane University. That contract provides for a systematic investigation of the relationships between personality, learning, and social variables. The project, under the direction of J. H. Rohrer, is a part of the research program of the Neuropsychiatry Branch, Bureau of Medicine and Surgery, U. S. Department of Navy. While this paper constitutes a technical report to ONR, the interpretations presented herein are those of the authors and do not represent, necessarily, those of the sponsoring agency, the Department of Navy.

could not be used for all observations. Although the slopes of the regression lines within the two groups were not significantly different and the regression line for the two means was linear, the regression coefficient for the data within the groups differed significantly from the regression coefficient for the group means.

Table 1 presents the mean and variance for the T scores for *O*s trained at 2 in., and 8 in. respectively. When *O*s who had been trained with a light that moved 2 in. were paired with *O*s trained with a light that moved 8 in., and these pairs were tested together for their judgments of autokinetic movements, the judgments converged to produce the X means also presented in Table 1. The variances for these scores are also presented. It may be noted that the variability of the judgments in the autokinetic session is considerably larger than the judgments in the more structured training session.

The group means for the autokinetic Y scores obtained one year after the X scores are also presented in Table 1 together with their variances. A reduction in the group means may be noted for both T groups. This reduction is statistically significant ($p < .01$) for the 8-in. training group but not for the 2-in. training group. Differences between sessions have been evaluated and are presented in Table 2.

In spite of the reduction in group means when retested after an interval of one year, there is still a high positive relationship between the norms established in sessions X and Y. This relationship is shown by the correlation coefficients in Table 3. The multiple-correlation coefficient, $R_{Y.XT}$, is approximately the same as the simple coefficient of correlation between first and second autokinetic scores.

In order to evaluate the relative contributions of X and T to the prediction of Y, the standard partial-regression coefficients were computed. The standard regression of Y on X, independent of T, is .739 and .600 for the 2-in. and 8-in. training groups respectively, while the standard regression of Y and T, independent of X, is .264 and −.008 for the 2-in. and 8-in. groups respectively. It is evident from a comparison of these coefficients that the contribution of T to the prediction of Y is negligible.

TABLE 1

MEANS AND VARIANCES

(N = 23 in both training sessions)

Session	2-in. Training		8-in. Training	
	Mean	Variance	Mean	Variance
T	3.08	6.08	8.74	.79
X	6.29	24.95	6.41	4.24
Y	5.50	12.54	5.09	5.06

TABLE 2

OBTAINED DIFFERENCES BETWEEN SESSIONS

(N = 23 in both training sessions)

Statistic	2-in. Training		8-in. Training	
	T-X	X-Y	T-X	X-Y
Mean diff.	3.21	0.79	2.33	1.32
t	2.675	0.6076	4.8724	2.0301
p	<.01	>.50	<.01	<.05

TABLE 3

INTERCORRELATION COEFFICIENTS

(N = 23 in both training sessions)

Coefficient	2-in. Training	8-in. Training
r_{XT}	.892	.595
r_{YT}	.693	.341
r_{XT}	.581	.582
$R_{Y.XT}$	.918	.598

DISCUSSION

In a previous study (4) it was demonstrated that the effects of training *O*s to judge the actual distance moved by a light *S*, generalized to an autokinetic light movement situation and thus influenced the magnitude of the perceived autokinetic movement. Stated in more formalistic terms, the habit strength of an *R* tendency built up to an *S* of real movement, generalized to an *S* situation in which the perceived movement was illusionary. The data presented in this study demonstrate that judgmental norms of perceived autokinetic movement, modified in the direction of conformance with judgments of a second person in an interpersonal (social) situation, tend to negate the effects of the earlier individual training. Moreover, this negation of training effects by social interaction persists in time, the effects of the social interaction manifesting itself in a "non-social" test situation after a period of one year. Again in more formalistic terms, the modification due to social interaction was in the intervening habit variable.

Kelman (5) has demonstrated that felt "success" leads to stability of norms established to perceived autokinetic movement and felt "failure" leads to norm modification. The terms "success" and "failure" are most meaningful when used within a theoretical framework employing the concept of reinforcement. Such a framework re-

quires the identification of "motives" or "drives" that, when reduced, strengthen the tendency to respond. The change made in our *S* situation, which produced the altered behavior, was the introduction of a second *O* who began giving reports of perceived movement that were at variance with the judgments given by the other member of the pair. The altered behavior that resulted brought into conformance the judgments of perceived movement reported by both members of the pair. It is hypothesized that this shift was activated by an acquired motive for social conformance.[2] The data argue for the shift being due to a drive variable, rather than due to a situational, or *S*, variable because the altered *R* persisted when the interpersonal interaction was not operating (session Y). Thus a stable *R* tendency was established to the light *S* in the 50-trial X session by the reduction of the conformity drive, and this *R* tendency was manifested in the subsequent individual Y session.

An attempt was made to explain the data by use of Bateson's (1) and Harlow's (3) notions of learning sets. The rather rapid rate of convergence noted in session X suggests that a set for conformance produced by the interpersonal cues could be involved. What is known about human socialization practices also suggests that sets to conform are one of the most likely by-products of socialization. Yet the results obtained in session Y suggest that the notion of learning sets is not sufficient to explain the results; the specificity and stability of the acquired *R* tendency suggests the theoretical need for an additional notion similar to what Harlow (3) has referred to as "Thorndikian bondage" resulting from drive reduction.

[2] Within the theoretical framework of Parsons *et al.* (6), this would be equivalent to their "need-dispositions to conform to expectations" that was felt by "ego." In the case of our *O*s, there was no evidence of overt conflict involved in the conforming behavior. They were not aware that they were conforming, yet all converged very rapidly, despite the presence of acquired *R* tendencies, established in the T session, that would cause them to tend *not* to conform.

SUMMARY

1. The stability of judgmental norms, established as a result of interpersonal interaction, was evaluated one year after they were established. The evaluation was made by testing the subjects individually.

2. It was shown that the norms, so established, were stable.

3. Data were presented which showed that social interaction resulted in a rapid modification of response tendencies established in a previous individual training session.

4. The results obtained were interpreted within a reinforcement theory framework.

REFERENCES

1. BATESON, G. Social planning and the concept of "deuterolearning." In T. M. Newcomb & E. L. Hartley (Eds.), *Readings in social psychology*. New York: Holt, 1947. Pp. 121–128.
2. BOVARD, E. W., JR. Social norms and the individual. *J. abnorm. soc. Psychol.*, 1948, **43**, 62–69.
3. HARLOW, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, **56**, 51–65.
4. HOFFMAN, E. L., SWANDER, D. V., BARON, S. H., & ROHRER, J. H. Generalization and exposure time as related to autokinetic movement. *J. exp. Psychol.*, 1953, **46**, 171–177.
5. KELMAN, H. C. Effects of success and failure on "suggestibility" in the autokinetic situation. *J. abnorm. soc. Psychol.*, 1950, **45**, 267–285.
6. PARSONS, T., & SHILS, E. A., (Eds.) *Toward a general theory of action*. Cambridge: Harvard Univer. Press, 1951. Pp. 159–189.
7. ROHRER, J. H., & HOFFMAN, E. L. An apparatus for studying the perception and judgment of light movement. *Amer. J. Psychol.*, 1954, **67**, 143–146.
8. SHERIF, M. A study of some social factors in perception. *Arch. Psychol.* New York, 1935, **27**, No. 187.

Received June 12, 1954. Prior publication.

STUDIES IN THE SCOPE AND METHOD OF "THE AUTHORITARIAN PERSONALITY." Edited by Richard Christie and Marie Jahoda. Glencoe, Illinois: The Free Press, 1954. Pp. 279. Price $4.50.

This is the second in a series of which *Studies in the Scope and Method of "The American Soldier"* was the first. Three of the book's six chapters are brief ones. Jahoda, in an editorial introduction, reconstructs some of the historical setting, both theoretical and methodological, in which the authors of *The Authoritatian Personality* began and continued their work. One of its unique features which she notes is its use of the methods of clinical psychology in conjunction with psychoanalytic theory "for other than clinical purposes, that is, for the exploration of political ideologies [as related to] personality structure." She also notes that none of the original authors was invited to reply to the criticisms (some of which, as will be shown, are rather harsh) in the volume under review here.

The other two short chapters, by E. A. Shils and H. D. Lasswell, respectively, present an interesting contrast. The latter refers hardly at all to the volume by Adorno *et al.* but, like it, is primarily concerned with the scientific problem of the contribution of personality variables to behavior in the public arena—specifically, in their contribution to political participation on the part of "the power-centered person." Shils's chapter, with its single-minded devotion to the charge that Adorno *et al.* were insufficiently aware of authoritarianism on the political "left," keeps its eye on the authors but not on their scientific problem. Lasswell's chapter, by contrast, is studded with relevant propositions, always interesting and generally testable.

The final chapter, by Frenkel-Brunswik, is appropriately titled "Further Explorations . . ." Though it contains an occasional explanation or defense of something in the original volume, it is in no sense an apologia. The largest portion of it is devoted to her own post-1950 investigations of phenomena in children similar to those originally reported for older subjects. For example, preliminary results from a set of experiments in perception "support the conjecture that, by and large, such tendencies as the quest for unqualified certainty, the rigid adherence to the given—be this an authority or a perceptual stimulus . . . tend to occur again and again within an individual in contexts seemingly far removed from each other."

In a final passage she raises the question whether or not the value biases of the authors of *The Authoritarian Personality* invalidated its conclusions. Among several "arguments" offered in defense of a negative answer is this interesting one: "From this angle we must welcome the fact that E. R. Jaensch has given us articulate description of the ideal personality type from the standpoint of Nazi ideology. The studies of Jaensch, reported in some detail above . . . have thus certainly been conducted under a bias opposite to ours, yet they were found to lead to the recognition of similar patterns albeit with a different evaluative slant."

The meat of the book is to be found in two long and critical chapters of which the first, by H. H. Hyman and P. B. Sheatsley, is a methodological one. Their conclusions, in the main convincing to this reviewer, are worth quoting at some length:

> The findings rest on an admittedly unrepresentative sample, from which large generalizations are incautiously drawn . . . .Weaknesses in the authors' handling of the quantitative data invariably work in favor of the assumptions, and the procedures are often inadmissible. . . . The questionnaire data are not validated save by their correspondence with the results of personal interviews with extreme groups and . . . the criterion for validation was not independently derived . . . no control at all was exercised over the important variable of formal education . . . .Statistical tables are presented, the relationships are noted, and they are then "explained" in terms of the authors' theory. But there is no safeguard on the interpretations, and they become mere statements of opinion . . . .Much of the qualitative analysis of respondents' ideology is almost completely unrestrained by any attachment to scientific method. The authors' propensity for psychodynamic explanations leads them to ignore the widespread distribution of certain sentiments in the American population . . . and to attribute them instead to psychodynamic processes unique in the ethnocentric individual.

These harsh judgments of the methodological trees do not, however, prevent the critics from taking an appreciative view of the research forest: "One feels intuitively that [the authors] have wisdom in their views and soundness in their conclusions, and it is sad that the acumen which inspired the project, the energy in executing a research task unparalleled in scope, and the intuitive power devoted to the appraisal of the results, were not matched by equal methodological skill."

In the other long chapter, "Authoritarianism Re-examined," R. Christie tackles five problems raised by the research reported in the original volume or by later researches inspired thereby. (*a*) The F-scale items "did not appear to capture authoritarianism *per se*. . . the postulated personality variables underlying their acceptance had tenuous but provocative support." (*b*) The F scale "did discriminate between individuals who dis-

played behavior which was authoritarian in interpersonal relations, rather than politically"; its discriminating power is "consistent" but not "exceptionally high." (*c*) The authors' contention that ethnocentrism is rather highly generalized is pretty well supported, and its relationship to "certain aspects of personality appears to be well established"; e.g., "those persons who view the world as threatening and unpredictable tend to be less tolerant of minority groups." (*d*) Various measures related to socioeconomic status show "a sizeable negative correlation" with F-scale scores; "these findings are interpreted as running counter to the general argument in *The Authoritatian Personality* which emphasized purely personality determinants . . . and discounted contemporary social influences." (*e*) With regard to "the hypothesized greater importance of early environment in contrast to adult experiences in the genesis of potentially fascistic personality . . . neither the original data nor subsequent findings are sufficient to prove or disprove the hypothesis."

Were it the present reviewer's function to offer a summary evaluation of *The Authoritarian Personality*, he could hardly do it more sagely than has Christie: "Both [its] strength and weakness lie in its basic assumptions which are rooted in psychoanalytic theory. Such an orientation has led to the uncovering of a host of data which in all likelihood would not have been discovered by investigators with differing theoretical viewpoints. Despite some methodological weaknesses in the original research, subsequent findings have been predominantly confirmatory."

It *is* this reviewer's function, however, to evaluate this volume of evaluations. Was it worth while to publish what might be termed a 275-page book review? The present verdict is a thrice-qualified affirmative. Qualification 1: Some chapters are carelessly edited and/or proofread. Qualification 2: All but the two long chapters might have found other publication outlets. Qualification 3: Only a work as important as *The Authoritarian Personality* could justify such an outsize review. In summary, the justification of the book under review lies in the importance of the book which *it* reviews, together with the carefully documented criticism of its two major chapters. Readers who enjoy fair, astute, detailed, and documented criticism of important psychological contributions will wish that there were more frequent opportunities for critiques like these.

THEODORE M. NEWCOMB
University of Michigan.

THE STRUCTURE OF PERSONALITY. By H. J. Eysenck. London: Methuen & Co. and New York: Wiley, 1953. Pp. 348. Price $5.75.

Some fifty years of endeavor on the part of several hundred psychologists are dissected and reported upon in this book. The literature was combed for empirical studies bearing on theories of personality structure, the proper statistical model for which, according to Dr. Eysenck, is *R* factor analysis, and more especially analysis into first- and second-order factors along Thurstone's well-known lines. The theory favored by Eysenck is a hierarchical construction of his own, according to which habitual modes of response become organized into traits, which in turn are arranged into a few all-embracing *types*. Traits and types correspond to first- and second-order factors, respectively. Traits are of the kind "shyness," "persistence," "rigidity"; types are groups of correlated traits, there being three such clusters, called introversion-extroversion, neuroticism, and psychoticism. The book is a disquisition on facts to support this hierarchical construction, drawn vicariously from studies on the analysis of questionnaires, objective behavior tests, physique, physiological measures, interests and attitudes, correlation between persons, and "trait" measurements.

Such a construction lends itself very well to Eysenck's stature in this field, and we may be sure that some penetrating analyses are undertaken in these pages, many sound conclusions as to fact drawn, and many interesting lines pursued. Eysenck takes a firm grip on his comb, pulling it where it will. Psychologists who have to deal with the practical problems of society, with the gross diagnosis of mental disorders, or with the selection of personnel in group testing situations and the like, will find a great deal indeed to please them in these chapters. The present reviewer does not feel competent to assay some of these facts, but where he is able to judge them in detail, as, for example, in the treatment of important studies by Jenkins and his collaborators (3), he is of the opinion that Eysenck's judgment is sound and constructive. The types themselves are perhaps badly named—introversion-extroversion in Eysenck's system is scarcely what Jung had in mind, and neuroticism ordinarily connotes a more specific set of conditions than Eysenck's second-order factor embraces. The chapter on correlation of persons is no worse for its inadequacies than others written by those who adopt the primacy of *R* methodology. One is sure, with such exceptions, that Eysenck has dealt fairly with facts, and that there is much substance to his construction.

Even so, a construction of this kind is not a theory in a genuine sense, and Eysenck's comb seems to have many broken teeth, so that much is left unteased on Psyche's shaggy brow. To take the lesser matter first; one may wonder whether or not important papers in the area, which happen to be of a theoretical nature, have been unwisely

passed over on the mere ground that they are not factual. Thus, it may be argued that a theoretical paper that relates perseveration to Pavlov's theories of inhibition (4), that is, substantially to conditions of true experimental neurosis, is worth more attention than most so-called factual work on the subject of perseveration. It is clear, also, that the armory of factor analysis does not end with first- and second-order factors, and that third and higher orders can be expected, or have been unearthed: what "theory," one wonders, will Eysenck offer for these? Since increase of complexity is embraced in the hierarchy, no doubt archetypes will correspond to third-order factors. Or, a third-order bipolar may perhaps correspond to the sexes, and fourth orders to the different races of mankind? These questions, it seems to this reviewer, lay bare the weakness of Eysenck's construction: clearly it is not a theory in any hypothetico-deductive sense, as Albino (1) has already indicated. Eysenck's construction is purely *ad hoc*, explaining nothing, and permitting of no pertinent hypotheses or deductions.

Here, it seems, is the essential difficulty. Nothing genuinely theoretical about personality, or even about scientific method, is to be found anywhere in these pages. This is the more astonishing because Eysenck professes above all else to be thoroughly scientific, and to be devoted to the hypothetico-deductive method. A theoretician, however, would have felt it essential to offer a theory for the types outlined in this book: several are possible, one, in particular, in relation to ego theory and defenses. Eysenck offers no such theory, of course, and would regard all of them, inevitably, as so much speculation. This is because as a factorist, he has grasped, above all else, the importance of operations in scientific work: for Eysenck a theory without actual tests and proofs is anathema. Most theories in personality study, Eysenck correctly sees, are without the essential operations. It may be suggested, however, that not all such theories are purely speculative. A little perspective might lead one to conclude that many are perhaps on the right lines, and that the factorist would do a really important job if he could provide the missing operations. The factor methods that Eysenck espouses so cleverly can in fact be placed at the service of the very theories that he rejects. Thus one must agree with Eysenck that matters of scientific proof are essential; one must agree, too, that factor methods can provide really cogent operations; but it is as well to remember that operations without a genuine theory in attendance are apt to become the worst kind of scientism (2), and it is from such a charge that the present reviewer would like to save Eysenck.

However, it will avail little to discuss this methodological difficulty. What is required from those of us who are factorists is considerable tolerance for all who have theoretical interests in personality studies. The necessary operations are not far away. And about the merits of our own efforts we might profit by remembering Queen Elizabeth I's outburst at Cambridge University nearly four centuries ago when her incomparable virtues were being extolled by the dons of her day: "Non est veritas," she blurted, "et utinam. . . ." So, too, for Eysenck and factor analysis. The time has surely come for a great deal of circumspection from all who are trying so hard to comb Psyche's hair. Otherwise we shall never find out whether, after all, she only wears a wig.

REFERENCES

1. ALBINO, R. C. Some criticism of the application of factor analysis to the study of personality. *Brit. J. Psychol. (Gen.)*, 1953, **44**, 164–168.
2. HAYEK, F. A. *The counter revolution of science: studies in the abuse of reason*. Glencoe, Ill.: Free Press, 1952.
3. LORR, M., & JENKINS, R. L. Patterns of maladjustment in children. *J. clin. Psychol.*, 1953, **9**, 16–19.
4. STEPHENSON, W. Studies in experimental psychiatry. *J. ment. Sci.*, 1932, **78**, 315–330.

WILLIAM STEPHENSON
University of Chicago

THE MAKING OF A SCIENTIST. By Anne Roe. New York: Dodd, Mead, 1953. Pp. ix + 244. Price $3.75.

This book by a psychologist is addressed primarily to nonpsychologists. Since the researches reported here are already available to psychologists in their professional journals, the interest of the volume for them will lie largely in the fact that it is an example of popular writing. The planning, the execution, the results, and the implications of the research are all here.

Most readers of this journal, doubtless, have a speaking acquaintance with Dr. Roe's project. She spent nearly four years making "clinical studies of vocational choice and performance in terms of life history and character structure." Sixty-four distinguished American-born research scientists in the age range 30 to 60 were her subjects. They were drawn in about equal numbers from the ranks of biologists, physicists, and social scientists (psychologists and anthropologists). A control group of competent but less distinguished scientists in the same fields was also studied by group methods. A control group of similarly distinguished nonscientists was not used. The results indicate that there are patterns in the life histories, intellectual abilities, and personality structure that are "more characteristic of scientists than they are of people at large, and some which are more characteristic of special kinds of scientists

than of other scientists." No differences were found between the more eminent and the less eminent physicists, but in the other two groups the more eminent "tended to have extremer degrees of the factors that characterized the men in the field."

The writing is clear and animated. The expository sections, especially those dealing with methodological problems and technical information, are models of simple, economical writing. The human interest value of the subject investigated is augmented by the chatty, autobiographical cast of the presentation itself, as well as by the inclusion of many personal experiences that were incidental to the conduct of the research. This added human interest, however, is had at a price: often the author, as a person, stands in front of the subject matter of the book, and thus distractingly competes for the reader's attention.

Dr. Roe tells us the several purposes she had in writing this volume. They were (*a*) to "enable me to review the experience as a whole," (*b*) "to give my subjects a more coherent and comprehensive account than the technical reports can give anyone but an expert in my field," (*c*) to "make it possible for others to share in a profoundly moving experience, and (*d*) to examine with me the meaning and value of science for the man who does it and for society" (p. 17). It is doubtful that writing this sort of book is essential to achieving the first purpose, and it is not likely that the author seriously believed it to be so. As to the second purpose, fully one fourth of the content of the book is not essential to achieving it; and much of this fourth, the reviewer believes, would annoy her distinguished subjects by its irrelevance. The third purpose, that of sharing a profoundly moving experience, is realized with a high degree of success. The fourth is really two purposes. One of these, the examination of the meaning of science for the man who does it, is achieved with distinction. The data are carefully evaluated, and the generalizations are circumspect. The other, the examination of the value and the meaning of science for society, is not attained. Here the responsible scientist yields to the beliefs and hopes of the socially conscious citizen-scientist. Properly speaking, Dr. Roe does not *examine* the meaning and value of science for society. Rather she testifies to a religious faith: ". . . the things that make good scientists are also the things that make good citizens" (p. 240). What is good for General Motors is good for the nation! According to the findings of her study, the things that make good scientists are experiences in the course of development which foster and make rewarding (*a*) independence in individual thinking, and (*b*) independent empirical investigation as a means of answering questions. Her data, however, do not indicate clearly or reliably that these two abilities and dispositions, which she found were characteristic of her subjects in the fields in which they were distinguished, are characteristic of them in other fields, nor do her data bear on the question of whether or not these characteristics are sufficient to make good citizens.

With her *faith* we cannot quarrel; indeed, many of us subscribe to it. But let us not embrace science and condemn dogma while at the same time we promulgate dogma in the name of science. It was not a scientist nor a psychoanalyst but a jurist, Mr. Justice Holmes, who advanced the opinion that the great difficulty with all thought is the inarticulate major premise.

Harvard University STANLEY G. ESTES

RACIAL AND CULTURAL MINORITIES: AN ANALYSIS OF PREJUDICE AND DISCRIMINATION. By George Eaton Simpson and J. Milton Yinger. New York: Harper and Brothers, 1953. Pp. ix + 773. Price $6.00.

In a few prefatory remarks the authors tell us why they wrote the book: the prejudice-discrimination melon, they say in effect, has been sliced a great many times and from a great many angles. This is a good thing. But the very proliferation of specialized studies may obscure the important truth that the melon is more than the sum of its slices; certainly it makes the task of synthesis increasingly difficult. Accordingly, it is time to re-examine the whole in the light of our accumulated knowledge about its component wedges. In the course of this re-examination, or in consequence of it, the authors hope to furnish the reader with guiding concepts which will make him ". . . less likely to adopt an oversimplified, one-factor view, and less likely to study one situation in isolation from the total society of which it is a part." They propose to do this in two ways. First, by ". . . relating the various analyses to each other and to a systematic group of principles that underlie them all," and secondly, by writing from the point of view that " . . .every proposition concerning intergroup relations should be . . . a part of the general principles being developed in social science." As it turns out (perhaps atypically) these objectives have a subtle but profound influence on the substance they preface.

There are three major sections. Part I, The Causes and Consequences of Prejudice and Discrimination (ten chapters, 315 pages), includes analyses of various types of majority-minority situations and of a number of "approaches" to the concept of race, a discussion of the personality and group-conflict functions of prejudice and of assorted consequences of prejudice and of discrimination, and a two-chapter "case study" of the sociology and social psychology of anti-Semitism. Part II, Minorities in the Social Structure: The

Institutional Patterns of Intergroup Relations (ten chapters, 324 pages), provides a one-chapter review of social class and caste notions and presents data on economic, political, legal, familial, religious, educational, and artistic characteristics of prominent minority groups in the United States. Part III, Prejudice Discrimination and Democratic Values (two chapters, 71 pages), comprises a program of strategies with which to reduce prejudice and discrimination by, respectively, changing the prejudiced person and changing situations conducive to prejudice. Remaining materials include a tabular appendix on the size of various minority groups in the United States, Brazil, and the Union of South Africa, a bibliography of 498 entries, and separate subject and name indexes.

In essence the book is a comprehensive review of historical and sociological fact and fancy regarding prejudice and discrimination and, as such, is potentially valuable for reference purposes. The authors' knowledge of their subject is impressive, their approach scholarly, even polymathic, and their prose learned but easy. For psychologists the second section particularly is likely to prove useful. It includes a wealth of factual material, admirably organized about specific behavioral areas, which is not otherwise available in a single work. For example, the four chapters in this section which are devoted to minorities and the American political and legal processes and to the education of minorities in the United States provide both a fascinating and a compelling setting for the recent Supreme Court decision on segregated schools and a remarkable demonstration of the institutionalization of attitudes toward minority groups. Unfortunately, the book's usefulness is impaired, perhaps vitiated, by several pervasive inadequacies. In the opinion of this reviewer, all are attributable to the familiar holistic bias implicit in the intent to "integrate" and to present a "total context." Two are illustrative.

First, the attempt to provide what one must suppose is a "multifactored" view and to consider the concepts and methods of a variety of disciplines leads to the inclusion of material which is of questionable relevance and/or dubious validity, and results in a diffuse and uneven treatment. What purports to be a critical evaluation turns out to be an uncritical admixture: demographic fact (Census data on the distribution of farms of nonwhite operators) and secondhand anecdote ("a Negro actress who travelled . . . through the South reported . . ."); hardheaded appraisal ("We have a measure of only the changes in responses to verbal tests, not of other behavioral changes"), and facile speculation ("for the minority artist, the creative process provides emotional release . . . [and] . . . social recognition"); historical erudition ("there were three major dispersals of Jews during antiquity"), and psychological naïveté (". . . prejudice . . . seized upon by an insecure person may result in . . . a rigidity of mind and compulsiveness in adjustment"); devotion to science ("to fail to take into account methodological problems is to give an impression of finality that no science . . . can accept") and methodological confusion ("In defining relations . . . with so many variables involved, one is bound to run into disagreement over what is more and what is less") are inextricably compounded. And while it may be true, for example, that "Reproduced, spurious, and bastard dialects tend to reinforce unfavorable racial stereotypes, [while] dialect poetry of the Sterling Brown type provides valuable insights into folk Negro life" (the reviewer and two colleagues were unable, in three tries, to separate the bastard and legitimate dialects), it is difficult to tell just where this observation fits. The task of sorting out the tangential and the suppositious falls to the reader.

Secondly, since the "general principles" which purportedly underlie the varied phenomena of minority-majority relations simply do not exist—the authors do not, at any rate, get around to stating them—the desired "synthesis" is accomplished by a kind of verbal stickum. Sometimes the stickum is a persuasive quotation from a gifted writer, for example, Lillian Smith or Richard Wright, and sometimes it is an equally persuasive passage by the gifted authors. In both cases, plausibility and fluency are apparently accepted as criteria of truth. In a discussion of the use of anti-prejudice propaganda, for example, the reader is admonished to heed the interpretive limits imposed by the available experimental evidence. These limits are restrictive; the experimental findings cannot easily be extrapolated to the pressing social problem of reducing prejudice. Accordingly, the discussion concludes with the quotation of a number of quite reasonable "principles that help one to understand the effectiveness of anti-prejudice propaganda." These principles have wide applicability and are wholly uncorroborated.

It is difficult to strike a balance between the book's principal asset—its truly exhaustive coverage of sociological and historical data and opinion—and its principal liability—its lack of selectivity. The answer undoubtedly depends in part on the level of sophistication of its potential readers. The reviewer is somewhat hesitant about recommending its use as a text, for example, for fear that undergraduates will be too easily taken in by its facile summaries and pseudosystematic bias. Those better qualified to pick its meat, however, will probably find it a useful refresher and reference source.

JOSEPH G. SMITH
Commission on Human Resources
and Advanced Training

SOCIAL PSYCHOLOGY AND INDIVIDUAL VALUES. By D. W. Harding. New York and London: Longmans, Green (Hutchinson's University Library), 1953. Pp. vii + 184. Price $2.40 (Text $1.80).

This book is concerned with what seems to be the central theme of social psychology, viz., the behavior of the individual in the group, or, as its author explicitly states it, with "the development of the social individual."

While he is fully aware of the opposing belief, Harding begins his study of the social individual with the widely held biological view that all animals, including man, are inherently social. Man's sociability is not, however, an instinct. Harding rejects the "herd instinct" of Trotter, the "gregarious propensity" of McDougall, and other similar modes of explanation. The error in the older views of the social nature of man, according to Harding, was their attribution of specificity to human sociability. What man wants most of the time, Harding believes, is "an awareness of friendly companionship in *other* activities, not specifically social ones" (p. 15), a knowledge that his activities are taking place in a congenial social context which stimulates a sense of being reciprocally sanctioned.

Although he prefers the biological explanation of an innate social disposition, the argument and the evidence which Harding submits are not crucial. As a matter of fact they might be used to support the contrary position equally well. To say, moreover, that this innate disposition requires a social context for its exercise can only mean operationally that it is an acquired habit in the first place. In the face of his own analysis the author would have been on firmer ground had he left the question of priority an open one.

Any attempt to account for the natural sociability of man inevitably runs into a collision with the stubborn fact of human pugnacity and hostility. Harding is as conscious of this problem as anyone, but his "solution" of it is unconvincing. In "integrative" behavior, he holds, conflict of aim is eliminated through the realization by the disputants that each looks upon the conflict from an exclusive perspective. By means of a special kind of cooperation, characterized by mutual influence in the use of means to the achievement of an end, they resolve their common difficulty. This is the wish, no doubt, of all reasonable men; but the realities of human conflict force one to describe its consummation as unduly optimistic.

Although man is naturally a social animal, his sociability, it has been said, is developed only in a social context. An explanation of his sociability is found in his early social development. The process of socialization, while long and varied, is summed up in the infant's attachment to his mother, who ministers to his needs and provides him with his basic satisfactions. Out of this attachment to the mother the infant develops a rudimentary sentiment, and, as his circle of interactions widens, he forms sentiments with respect to other individuals and their behavior, the final product of which is his own personality.

Unlike many psychologists, Harding is not afraid to tackle the problem of the moral development of man. In a chapter on "The Individual's Morality" he deals with moral standards without moralizing. Being social beings, men place a high value on consideration for others. Whence comes this need to do right even when detection of doing wrong is unlikely? There are two reasons for this, one individual, the other social. "A breach of the moral code that we share with other people," Harding writes, "automatically deprives us of our sense of having group support in what we do: we know that our group now sanctions us not for what we are but for what we pretend to be" (p. 54).

The individual explanation is based on the holistic character of man. This consists in "the central effort of a living being to remain an integrated whole and avoid tensions that result from actions conflicting with the stable features of his personality, such as his sentiments and the established hierarchy of his values" (p. 54). Coherent moral behavior is not derived from membership in a social group but is a characteristic of man's biological make-up.

In a period of the development of social psychology when the Group, like the King of old, can do no wrong, it is gratifying to find in this book an examination of the group's fitness for its members. Adjustment and conformity are here put in a balanced perspective. An important psychological problem is to examine the extent to which the group permits and to what extent it stifles the individual's satisfaction of private needs and desires and his ability to adhere to values that other members of the group will not support. A group whose culture fails to bring out an individual's possibilities, or one which, through rejection or isolation, handicaps its gifted members, is inadequate to their needs. In this connection the author rightly criticizes the trend of much social psychology and psychotherapy toward a passive and uncritical acceptance of the standards of the dominant social group.

In the chapter on "The Social Meaning of Normality" we find familiar material, but it bears repetition. Like most psychologists, Harding respects the statistical conception of normality, but he is not beguiled by its quantitative formulation. Neither does he go the whole way with those cultural relativists who would define normality and abnormality wholly in terms of cultural norms. A view that will do justice to the richness and com-

plexity of human behavior, he shows, must also consider the degree of integration or wholeness of the individual. Much ill health, he says, expresses "the mind's failure to achieve or maintain integration"; and an individual is "maladjusted" to the extent that his mind is "divided against itself and failing to maintain its wholeness" (p. 118).

There are other important topics in this book that deserve consideration but which must be omitted in this brief review. Harding's analysis of leadership, a topic yet to be adequately treated by social psychologists, deserves the reader's careful attention. The term "leader," he suggests, "might profitably be kept for one who guides and coordinates the efforts of a number of people (or at least one other person) while they perform some activity" (p. 107). A chapter each on "Innovators" and "The Average and the Excellent" round out the book and suggest a measure of the sensitivity of the author to human problems.

The American reader of this British book who gauges its adequacy by textbook standards will miss a number of conventional topics. On occasion he may even become impatient with some of its philosophical overtones. These are on close examination, however, the astute observations by a sensitive mind of some of the broad implications of empirical research in social psychology. While the author is bound by the facts, he is not their slave. The result is a fine combination of scientific truth, practical wisdom, and literary excellence.

HUBERT BONNER
Columbia University.

NEW BOOKS RECEIVED

ABRAMSON, H. A. (Ed.) *Problems of consciousness.* New York: Josiah Macy, 1954. Pp. 177. Price $3.25.

ARNOLD, MAGDA B., & GASSON, J. A. *The human person.* New York: Ronald, 1954. Pp. x + 593. Price $5.75.

BAKER, JAMES, JR. *The exteriorization of the mental body.* New York: William-Frederick Press, 1954. Pp. 32. Price $1.50.

BARRELL, JOSEPH. *A philosophical study of the human mind.* New York: Philosophical Library, 1954. Pp. xi + 575. Price $6.00.

BOSSARD, J. H. S. *The sociology of child development.* (Rev. Ed.). New York: Harpers, 1954. Pp. xi + 788. Price $6.00.

BRAATOY, TRYGVE. *Fundamentals of psychoanalytic technique.* New York: Wiley, 1954. Pp. xi + 404. Price $6.00.

BRAY, D. W. *Issues in the study of talent.* New York: King's Crown Press, 1954. Pp. xi + 65. Price $2.00.

BURMEISTER, EVA. *Roofs for the family.* New York: Columbia Univer. Press, 1954. Pp. 203. Price $3.25.

BURT, SIR CYRIL. *The causes and treatment of backwardness.* New York: Philosophical Library, 1954. Pp. 128. Price $3.75.

CAPLOW, THEODORE. *The sociology of work.* Minneapolis: Univer. of Minnesota Press, 1954. Pp. viii + 330. Price $5.00.

CHINOY, ELY. *Sociological perspective: basic concepts and their application.* New York: Doubleday, 1954. Pp. vi + 58. Price $.85.

COMMINS, W. D., & FAGIN, BARRY. *Principles of education psychology.* (2nd Ed.) New York: Ronald, 1954. Pp. xvi + 795. Price $5.75.

CRONBACH, LEE J. *Educational psychology.* New York: Harcourt, Brace, 1954. Pp. xxvii + 628. Price $7.50.

DONIGER, SIMON. (Ed.) *Religion and human behavior.* New York: Association Press, 1954. Pp. xxii + 233. Price $3.00.

ENGLISH, O. P., & FINCH, S. M. *Introduction to psychiatry.* New York: Norton, 1954. Pp. viii + 621. Price $7.00.

FABIAN, ROBERT. *London after dark.* New York: British Book Center, 1954. Pp. 237. Price $3.50.

FINK, H. K. *Long journey.* New York: Julian Press, 1954. Pp. xviii + 298.

FINK, H. K. *Mind and performance.* New York: Vantage Press, 1954. Pp. xi + 113. Price $3.00.

FRASER, D. C. *A psychological glossary.* Cambridge, England: Heffer & Sons, 1954. No pages numbered. Price 3s6d.

FROSCH, JOHN. (Ed.) *The annual survey of psychoanalysis.* Vol. II. New York: International Universities Press, 1954. Pp. xiv + 724 Price $10.00.

GALDSTON, IAGO. (Ed.) *Beyond the germ theory.* New York: Health Education Council, 1954. Pp. viii + 182. Price $4.00.

GEDDES, D. P. (Ed.) *An analysis of the Kinsey reports.* New York: Dutton, 1954. Pp. x + 319. Price $3.50.

GILL, MERTON, NEWMAN, RICHARD, & REDLICH, F. C. *The initial interview in psychiatric practice.* New York: International Universities Press, 1954. Pp. 423. Price $6.00. plus phonograph records $4.60.

GORER, GEOFFREY. *The life and ideas of the Marquis de Sade.* New York: British Book Centre, 1954. PP. 244. Price $3.50.

GUTKIND, E. A. *Community and environment.* New York: Philosophical Library, 1954. Pp. xiii + 81. Price $3.75.

HADFIELD, J. A. *Dreams and nightmares.* Baltimore: Penguin Books, 1954. Pp. x + 243. Price $ .65.

HEATH, A. G., *et al. Studies in schizophrenia.* Cambridge: Harvard Univer. Press, 1954. Pp. xiv + 619. Price $8.50.

HODGSON, K. W. *The deaf and their problems.* New York: Philosophical Library, 1954. Pp. xx + 348. Price $6.00.

HOFSTÄTTER, P. R. *Einführung in die Sozialpsychologie.* Stuttgart-Wien: Humboldt-Verlag, 1954. Pp. 535. Price: DM 12.50.

ITTELSON, W. H., & Cantril, Hadley. *Perception: a transactional approach.* New York: Doubleday, 1954. Pp. ix + 33. Price $ .85.

JERSILD, A. T. *Child psychology.* (4th Ed.) New York: Prentice-Hall, 1954. Pp. xii + 676. Price $6.00.

JUNG, C. G. (Trans. by R. F. C. Hull.) *The practice of psychotherapy.* Vol. XVI of Collected Works of Jung. New York: Pantheon (Bollingen Series XX), 1954. Pp. xi + 377. Price $4.50.

KATZ, DANIEL. (Ed.) *Public opinion and propaganda.* New York: Dryden, 1954. Pp. xx + 779. Price $6.25.

KLINEBERG, OTTO. *Social psychology.* (Rev. Ed.) New York: Holt, 1954. Pp. xiii + 578. Price $5.25.

KOHLER, CLAUDE. *Les déficiences intellectuelles chez l'enfant.* Paris: Presses Universitaires de France, 1954. Pp. 220. Price not supplied.

LANDER, BERNARD. *Towards an understanding of juvenile delinquency.* New York: Columbia Univer. Press, 1954. Pp. xv + 143. Price $3.00.

Learning theory, personality theory and clinical re-

search. The Kentucky Symposium. New York: Wiley, 1954. Pp. ix + 164. Price $3.50.

LEWIS, N. D. C., & ENGLE, BERNICE. (Eds.) *Wartime psychiatry: a compendium of international literature*. New York: Oxford Univer. Press, 1954. Pp. vi + 952. Price $15.00.

LINDGREN, H. C. *Mental health in education*. New York: Holt, 1954. Pp. xiii + 561. Price $4.75.

MCGILL, V. J. *Emotions and reason*. Springfield, Ill.: Charles C Thomas, 1954. Pp. xiii + 122. Price $3.25.

MARTIN, J. B. *Break down the walls*. New York: Ballantine Books, 1954. Pp. 310.

MYKLEBUST, H. R. *Auditory disorders in children*. New York: Grune & Stratton, 1954. Pp. xiii + 367. Price $6.00.

NEUMANN, ERICH (Trans. by R. F. C. Hull.) *The origins and history of consciousness*. New York: Pantheon (Bollingen Series XLII), 1954. Pp. xxiv + 493. Price $5.00.

ODLUM, DORIS M. *Psychology, the nurse, and the patient*. New York: Philosophical Library, 1954. Pp. 168. Price $4.75.

PARSONS, TALCOTT. *Essays in sociological theory*. (Rev. Ed.) Glencoe, Ill.: Free Press, 1954. Pp. 459. Price $6.00.

PEARSEN, G. H. J. *Psychoanalysis and the education of the child*. New York: Norton, 1954. Pp. x + 357. Price $5.00.

PENNINGTON, L. A., & BERG, I. A. *An introduction to clinical psychology*. (2nd Ed.) New York: Ronald, 1954. Pp. vii + 709. Price $6.50.

PEPINSKY, H. B., & PEPINSKY, PAULINE N. *Counseling: theory and practice*. New York: Ronald, 1954. Pp. viii + 307. Price $4.50.

REMMERS, H. H.: *Introduction to opinion and attitude measurement*. New York: Harpers, 1954. Pp. viii + 437. Price $5.00.

RIESMAN, DAVID. *Individualism reconsidered*. Glencoe, Ill.: Free Press, 1954. Pp. 529. Price $6.00.

RUNES, D. D. *Letters to my daughter*. New York: Philosophical Library, 1954. Pp. 131. Price $2.50.

RYAN, T. A., & SMITH, PATRICIA C. *Principles of industrial psychology*. New York: Ronald, 1954. Pp. xiv + 534. Price $5.50.

SARASON, S. B. *The clinical interaction*. New York: Harpers, 1954. Pp. x + 425. Price $5.00.

SCHANCK. R. L. *The permanent revolution in science*. New York: Philosophical Library, 1954. Pp. xvi + 112. Price $3.00.

Selected writings of de Sade. (Selected and translated by Leonard de Saint-Yves) New York: British Book Centre, 1954. Pp. 306. Price $6.75.

SMITH, J. A. *Alcoholism*. Philadelphia: Lippincott, 1954. Pp. vii + 72. Price $3.00.

SOLOMON, J. C. *A synthesis of human behavior*. New York: Grune & Stratton, 1954. Pp. xii + 265. Price $5.50.

SULLIVAN, H. S. *The psychiatric interview*. New York: Norton, 1954. Pp. xxiii + 246. Price $4.50.

THE TULANE DEPARTMENT OF PSYCHIATRY AND NEUROLOGY. *Studies in schizophrenia: a multi-disciplinary approach to mind-brain relationships*. Cambridge: Harvard Univer. Press, 1954. Pp. xiv + 619. Price $8.50.

THORPE, L. P., & SCHMULLER, A. M. *Contemporary theories of learning*. New York: Ronald, 1954. Pp. viii + 480. Price $5.50.

WOLBERG, L. R. *Technique of psychotherapy*. New York: Grune & Stratton, 1954. Pp. xiv + 869. Price $14.75.

WOLSTEIN, BENJAMIN. *Transference: its meaning and function*. New York: Grune & Stratton, 1954. Pp. xiii + 206. Price $5.00.

WULFECK, J. W., & BENNETT, E. M. *The language of dynamic psychology as related to motivation research*. New York: McGraw-Hill, 1954. Pp. 111. Price $4.00.

ZILBOORG, GREGORY. *The psychology of the criminal act and punishment*. New York: Harcourt, Brace, 1954. Pp. xi + 141. Price $3.50.

American Sociological Review

Official Journal of the American Sociological Society

announces

SMALL GROUP RESEARCH

As the December 1954 issue

edited by Fred L. Strodtbeck

The Coalition Pattern in Three Person Groups, *by Theodore M. Mills*

An Experiment on Conformity to Group Norms Where Conformity is Detrimental to Group Achievement, *by Harold H. Kelley and Martin M. Shapiro*

The Relation Between Group Members' Self-Confidence and Their Reactions to Groups Pressures to Uniformity, *by Godfrey M. Hochbaum*

Deviation and Rejection: An Experimental Replication, *by Richard M. Emerson*

A Study of Group Action, *by David G. Hays and Robert R. Bush*

Search Behavior in Individual and Group Problem Solving, *by Omar Khayyam Moore and Scarvia B. Anderson*

Interpersonal Orientations in Small Groups: A Consideration of the Questionnaire Approach, *by Matilda White Riley, Richard Cohn, Jackson Toby, and John W. Riley, Jr.*

The Cash Posters: A Study of a Group of Working Girls, *by George C. Homans*

Group Norms and the Active Minority, *by James G. March*

The Behavior of Small Groups under the Stress, Conditions of "Survival," *by E. Paul Torrance*

Some Findings Relevant to the Great Man Theory of Leadership, *by Edgar F. Borgatta and Arthur S. Couch*

Toward Integrating Field Work and Laboratory in Small Group Research, *by Muzafer Sherif*

Planning an Observation Room and Group Laboratory, *by Robert F. Bales and Ned A. Flanders*

$1.25 per copy

SPECIAL 10% DISCOUNT

On orders of 5 or more copies (for use in teaching)

AMERICAN SOCIOLOGICAL SOCIETY

New York University

Washington Square, New York 3, N. Y.

Journal of Abnormal and Social Psychology

AVAILABLE BACK ISSUES

July 1, 1954

Year	Volume	Available Numbers						Price per Number	Price per Volume
1906-07	1	—	—	—	—	—	—	—	—
1907-08	2	—	—	3	—	—	6	$2.00	—
1908-09	3	—	—	—	—	—	—	—	—
1909-10	4	—	—	—	—	—	—	—	—
1910-11	5	—	2	—	—	—	6	$2.00	—
1911-12	6	—	—	—	—	—	—	—	—
1912-13	7	—	—	—	—	—	—	—	—
1913-14	8	—	—	—	—	—	—	—	—
1914-15	9	—	—	—	—	—	—	—	—
1915-16	10	—	—	—	—	—	—	—	—
1916-17	11	—	—	—	—	—	—	—	—
1917-18	12	—	—	—	—	—	—	—	—
1918-19	13	1	2	—	—	—	—	$2.00	—
1919-20	14	—	—	3	4	5	6	$2.00	—
1920-21	15	1	—	—	4	—	—	$2.00	—
1921-22	16	—	—	—	—	—	—	—	—
		Apr.	July	Oct.	Jan.				
1922-23	17	1	—	—	—			$2.00	—
1923-24	18	—	—	—	—			—	—
1924-25	19	1	2	3	4			$2.00	$8.00
1925-26	20	—	—	—	—			—	—
1926-27	21	—	—	—	—			—	—
1927-28	22	—	—	—	4			$2.00	—
1928-29	23	—	—	—	4			$2.00	—
1929-30	24	—	2	3	4			$2.00	—
1930-31	25	1	2	3	4			$2.00	$8.00
1931-32	26	1	2	3	4			$2.00	$8.00
1932-33	27	1	2	3	4			$2.00	$8.00
1933-34	28	1	2	3	4			$2.00	$8.00
1934-35	29	1	2	3	4			$2.00	$8.00
1935-36	30	1	2	3	4			$2.00	$8.00
1936-37	31	1	2	3	4			$2.00	$8.00
		Jan.	Apr.	July	Oct.				
1937	32	1	2	3	&	4		$2.00	$8.00
1938	33	1	2	3		4		$2.00	$8.00
1939	34	1	—	3		4		$2.00	$8.00
1940	35	—	—	—	—			—	—
1941	36	—	—	—	—			—	—
1942	37	1	2	3		4		$2.00	$8.00
1943	38	1	2	3		4		$2.00	$8.00
1944	39	1	2	3		4		$2.00	$8.00
1945	40	1	2	3		4		$2.00	$8.00
1946	41	1	2	3		4		$2.00	$8.00
1947	42	1	2	—		4		$2.00	—
1948	43	1	2	—		4		$2.00	—
1949	44	1	2	3		4		$2.00	$8.00
1950	45	1	2	3		4		$2.00	$8.00
1951	46	1	2	3		4		$2.00	$8.00
1952	47	1	2	3		4		$2.00	$8.00
1953	48	1	2	3		4		$2.00	$8.00
1954	49	By subscription $7.00						$2.00	—

Postage prepaid on U. S. orders. Add $.50 per volume on foreign orders. All stock subject to prior sale.

The American Psychological Association gives the following discounts:

10% on orders of $50.00 or over for any one journal
20% on orders of $100.00 or over for any one journal
30% on orders of $150.00 or over for any one journal

Orders for available back issues and for current subscriptions should be addressed to:

American Psychological Association

1333 Sixteenth Street N. W. Washington 6, D. C.

INTERNATIONAL SOCIAL SCIENCE BULLETIN

published quarterly by the United Nations Educational, Scientific and Cultural Organization, 19 Avenue Kleber, Paris 16e.

Just issued: Vol. VI, No. 1

PART I: THE SECOND WORLD CONGRESS OF SOCIOLOGY, LIEGE 1953

PART II: ORGANIZATION IN THE SOCIAL SCIENCES, REVIEWS AND ANNOUNCEMENTS

REVIEW OF DOCUMENTS, PERIODICALS AND BOOKS

NEWS AND ANNOUNCEMENTS

OPEN FORUM: THE INDIVIDUAL AND SOCIETY, *by M. Ginsberg*

Annual Subscription: $3.50
Single copy: $1.00

Send subscriptions to:

COLUMBIA UNIVERSITY PRESS
2960 Broadway New York 27, N. Y.

Again Available

SYMPOSIUM: PSYCHOANALYSIS AS SEEN BY ANALYZED PSYCHOLOGISTS

This well-known symposium, which first appeared in the *Journal of Abnormal and Social Psychology* in 1940, has now been reprinted. The contributors are Edwin G. Boring, Carney Landis, J. F. Brown, Raymond R. Willoughby, Percival M. Symonds, Henry A. Murray, Else Frenkel-Brunswik, David Shakow, and Donald V. McGranahan. Comments on the symposium are made by Hanns Sachs and Franz Alexander. 160 pages.

Price, $2.00

Order from

American Psychological Association

1333 Sixteenth Street N. W. Washington 6, D. C.

— PASTORAL PSYCHOLOGY —

A monthly devoted to the practical synthesis of the principles and dynamics of clinical psychology, dynamic psychiatry, and psychiatric social work with spiritual and religious values

EDITORIAL ADVISORY BOARD

REV. OREN H. BAKER, Dean
The Colgate-Rochester Divinity School

DANIEL BLAIN, M.D.
Medical Director, American Psychiatric Assn.

REV. JOHN S. BONNELL
Fifth Avenue Presbyterian Church, N. Y. C.

REV. ERNEST E. BRUDER
St. Elizabeth's Hospital, Washington, D. C.

REV. ROY A. BURKHART
First Community Church, Columbus, Ohio

REV. RUSSELL L. DICKS
The Divinity School, Duke University

REV. DAVID D. EITZEN
School of Religion, Univ. of Southern California

REV. ROLLIN J. FAIRBANKS
Episcopal Theological School, Cambridge, Mass.

LAWRENCE K. FRANK
New York City

REV. MYRON T. HOPPER, Dean
College of the Bible, Lexington, Ky.

REV. PAUL E. JOHNSON
Boston University, School of Theology

REV. HALFORD E. LUCCOCK
Yale Divinity School

ROLLO MAY, Ph.D.
Consulting Psychologist, N. Y. C.

WILLIAM C. MENNINGER, M.D.
The Menninger Foundation

JOHN A. P. MILLET, M.D.
American Rehabilitation Committee

REV. WAYNE E. OATES
The Southern Baptist Theological Seminary

REV. DAVID E. ROBERTS
Union Theological Seminary

CARL R. ROGERS, Ph.D.
University of Chicago

REV. PAUL TILLICH
Union Theological Seminary

REV. GRANGER WESTBERG
Lutheran Advisory Board

REV. CARROLL A. WISE
Garrett Biblical Institute

LUTHER E. WOODWARD, Ph.D.
New York State Dept. of Mental Hygiene

REV. SEWARD HILTNER, *Pastoral Consultant* — SIMON DONIGER, Ph.D., *Editor*

Here are a few of our important articles:

RELIGIOUS APPLICATION OF PSYCHIATRY ..... Karl Menninger, M.D.
RELIGION, PSYCHOTHERAPY, SELFHOOD ..... Rollo May, Ph.D.
THEOLOGY OF SEX RELATIONS ..... Reuel L. Howe, D.D.
THE STRESSES AND STRAINS OF MARRIAGE ..... Erich Lindemann, M.D.
A BASIC ORIENTATION TO COUNSELING ..... Carl R. Rogers, Ph.D.
DYNAMICS OF PERSONALITY DEVELOPMENT ..... Franz Alexander, M.D.
THE UNLOVING PERSONALITY AND THE RELIGION OF LOVE ..... Bonaro W. Overstreet
THE DIAGNOSTIC USE OF THE BIBLE ..... Wayne E. Oates, D.D.
THE ROOTS OF RELIGION ..... Gordon W. Allport, Ph.D.
THE PASTOR AND THE CLINICAL PSYCHOLOGIST ..... Molly Harrower, Ph.D.
WHAT IS HAPPENING TO THE AMERICAN FAMILY? ..... Margaret Mead, Ph.D.
THE ART OF DREAM INTERPRETATION ..... Erich Fromm, Ph.D.
ONE WAY TO PERSONALITY ..... Lawrence K. Frank
THE ACT OF SURRENDER IN THE TREATMENT OF THE ALCOHOLIC ..... Harry M. Tiebout, M.D.
GROUP THERAPY IN SEXUAL MALADJUSTMENT ..... Abraham Stone, M.D.
SOCIAL CONTEXT OF PERSONALITY ..... Dean Liston Pope, D.D.
SIN AND SALVATION ..... Dean Willard L. Sperry, D.D.

PASTORAL PSYCHOLOGY, Great Neck, N. Y.

I enclose ☐ $5 for a one-year subscription

☐ $11 for a three-year subscription

(Canadian and foreign subscriptions 50¢ per year additional)

NAME .....

ADDRESS .....

CITY ..... ZONE ..... STATE .....

PSYCHOMOTOR ASPECTS
of
MENTAL DISEASE

An Experimental Study
By H. E. KING, Ph.D.

THE RESEARCH here reported tested the theory that body movements have a definite bio-psychologic character—that they are rooted in and made possible by biologic mechanisms while they act upon the external world for the satisfaction of needs and for survival. The performance of five groups of abnormal subjects, ranged in the order of the severity of their behavior disorders, was compared with that of a group of normal subjects of comparable age and sex. The results show a definite correlation between the disordered mental states of the patients studied and the disorganization of their body movements.

THE BOOK thus reveals a unique demonstration of a somatic concomitant of disordered mental states. It makes available a quantified technique applicable both to the diagnosis of mental disease and to the measurement of the patient's momentary status. It provides a theoretical framework for further development of both research and clinical test methods. And it includes the only broad review of the experimental literature related to this topic. It is of obvious importance to psychiatrists, psychologists, neurologists, and neurophysiologists.

200 pages, 31 figures, 21 tables, $5.50

THIS NEW WORK is A COMMONWEALTH FUND BOOK

Buy it at your bookstore, or write:
DEPT. CF-25
HARVARD UNIVERSITY PRESS
Cambridge 38 Massachusetts

Four Important Psychology Texts

Dynamic and Abnormal Psychology

W. S. TAYLOR, *Smith College*

"Well organized and written from a rich background. A highly successful [illegible] date."—Paul E. Fraser, Florida State University

"An excellent discussion of the basic psychological phenomena necessary for an understanding of both normal and abnormal behavior."—W. E. Hunt, [illegible] College

Social Psychology • An Interdisciplinary Approach

HERBERT BONNER, *Ohio Wesleyan University*

"The interdisciplinary aspect of this book answers a long-felt need. It is a real contribution."—Gordon W. Blackwell, University of North Carolina

"This is a well-written, highly comprehensive text in social psychology which reaches out into anthropology and sociology."—Dwight W. Chapman, Vassar College

A Laboratory Manual for Social Psychology

WILBERT S. RAY, *Research of Trinity College, Hartford, Connecticut*

"Delighted with Ray's *Manual*. Long needed for thorough training in social psychology."—C. R. Carpenter, Pennsylvania State College

"The field research studies, added to the laboratory experiments, raise the student into the practical aspects of research."—Christian O. Weber, [illegible] College

Psychology of Personal and Social Adjustment

HENRY CLAY LINDGREN, *San Francisco State College*

"A sound, helpful, and up-to-date book."—C. R. Rogers, Univ. of Chicago

"I am pleased that the author has presented us with this comprehensive coverage of problems of mental hygiene."—C. M. Gjerde, San Diego State College

AMERICAN BOOK COMPANY

COLLEGE DIVISION • 55 FIFTH AVENUE, NEW YORK 3, N. Y.

CINCINNATI • CHICAGO • ATLANTA • DALLAS • SAN FRANCISCO